

TruSight Oncology 500: Enabling Comprehensive Genomic Profiling and Biomarker Reporting with Targeted Sequencing

Chen Zhao^{1,#}, Tingting Jiang¹, Jin Hyun Ju¹, Shile Zhang¹, Jenhan Tao¹, Yao Fu¹, Jenn Lococo¹, Janel Dockter¹, Traci Pawlowski¹ and Sven Bilke¹

¹Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA

[#]Corresponding author: czhao@illumina.com

Keywords: TruSight Oncology 500, tumor mutational burden, microsatellite instability, targeted sequencing, variant calling

Abstract

Background: As knowledge of mechanisms that drive the development of cancer grows, there has been corresponding growth in therapies specific to a mechanism. While these therapies show improvements in patient outcomes, they can be expensive and are effective only for a subset of patients. These treatments drive interest in research focused on the assignment of cancer therapies based on aberrations in individual genes or biomarkers that assess the broader mutational landscape, including microsatellite instability (MSI) and tumor mutational burden (TMB).

Methods: Here we describe the TruSight™ Oncology 500 (TSO500; Research Use Only) bioinformatics workflow. This tumor-only approach leverages the next-generation sequencing-based assay TSO500 to enable high fidelity determination of DNA variants across 523 cancer-relevant genes, as well as MSI status and TMB in formalin-fixed paraffin-embedded (FFPE) samples.

Results: The TSO500 bioinformatic workflow integrates unique molecular identifier (UMI)-based error correction and a dual approach variant filtering strategy that combines statistical modeling of error rates and database annotations to achieve detection of variants with allele frequency approaching 5% with 99.9998% per base specificity and 99% sensitivity in FFPE samples representing a variety of tumor types. TMB determined using the tumor-only workflow of TSO500 correlated well with tumor-normal (N =170, adjusted $R^2=0.9945$) and whole-exome sequencing (N=108, adjusted $R^2=0.933$). Similarly, MSI status determined by TSO500 showed agreement (N=106, 98% agreement) with a MSI-PCR assay.

Conclusion: TSO500 is an accurate tumor-only workflow that enables researchers to systematically characterize tumors and identify the next generation of clinical biomarkers.

Background

It is well established that cancer is a diverse group of diseases resulting from genetic changes that facilitate aberrant and uncontrolled cellular proliferation. Indeed, as the number of known cancer-related genes and understanding of their mode of action grows, there has been a progressive shift towards assigning therapies based on the gene variants present in the tumor of a patient. The growing understanding of the role of genetics in cancer has been enabled by advancements in next-generation sequencing (NGS) technologies that have made the assessment of tumor-related gene alterations easier and more economically feasible. As the number of gene-based therapies and the evidence supporting their increased efficacy grows, there is rising demand for solutions that enable the development of accurate biomarkers that can identify patients most likely to benefit from these therapies in a personalized manner. Biomarkers range from variants in individual genes (eg. BRAF V600E, EGFR) (1-4) as well as others that take into account the broader mutational landscape of a tumor.

Detecting genetic variants in tumors can be difficult due to technical artifacts associated with FFPE samples, which are routinely used in translational research settings. The FFPE process can introduce DNA damage such as deamination (5). Additionally, the amount of tumor content may be a small fraction of the total FFPE sample, and so variants of interest may occur at low frequencies. The detection of somatic variants is further complicated by the high frequency of germline variants (hundreds of mutations per megabase) relative to somatic variants (tens of mutations per megabase) and technical artifacts (over one thousand per megabase) (5-8). While germline variants can be identified by sequencing matched tumor/normal samples, this considerably increases the cost of using NGS for tumor profiling, which remains an important consideration for translational researchers (9).

Microsatellite instability (MSI) status determined using immunohistochemistry (IHC) and polymerase chain reaction (PCR), was the first tissue site-agnostic biomarker approved by the United States Food and Drug Administration for use in selecting patients for immunotherapy. (10,11). Microsatellites are small regions, tens to hundreds of bases in length, of short repeating DNA motifs (1-9 nucleotides). Microsatellite instability is an increase or decrease in the number of these repeats resulting from a mutation in one of the mismatch repair genes - MLH1, MSH2,

MSH6 and PMS2 (12,13). Le et al (14,15) were the first to report that mismatch repair-deficient colorectal cancer patients, identified by MSI status, responded better to checkpoint inhibitors. Immunohistochemistry and PCR are established technologies for assessing MSI status (16,17). PCR approaches typically measures the length of five loci, which were first recommended by the Bethesda panel, in the tumor as well as matched normal tissue (16). IHC assays measure the protein level of the mismatch repair genes. Emerging NGS methods for determining MSI status assess many more microsatellite loci than PCR, which improves sensitivity and does not require a matched normal sample (18).

The prevalence of MSI varies between cancers and there are some cancer types that are mostly microsatellite stable, indicating a need for additional biomarkers to complement MSI status (19). Tumor mutational burden (TMB), which reflects the number of tumor mutations per megabase (Mb) of DNA, has been proposed as an additional biomarker. Studies have shown a correlation between higher TMB and the effectiveness of checkpoint inhibitor immunotherapies.(20-22) The current gold standard for determining TMB is whole exome sequencing (WES), but uptake is limited by the cost and long turn-around-time (23). Thus, there is significant interest in the use of targeted panels to determine TMB. There are a number of key factors that influence TMB calculation: panel size and composition, read depth and coverage, assay limit of detection (LoD), and the accuracy of the variant calls (24). Studies have shown that panels over 1 Mb are generally suitable for TMB determination (24,25). With smaller panels, the confidence intervals are too large at lower TMB values (0-30 mutations/Mb) to allow clear distinction between TMB low and TMB high samples. At present there is no established consensus around thresholds for TMB-based therapy selection. Further, the process by which TMB is calculated is not standardized and the bioinformatics behind different commercial assays are a veritable black box.

Here, we describe the bioinformatics workflow of TruSight™ Oncology 500 (TSO500), a pan-cancer NGS assay covering 523 cancer-relevant genes (1.94 Mb) that enables determination of DNA and RNA somatic variants as well as the immuno-oncology relevant biomarkers, MSI and TMB, from a FFPE tumor sample only. The TSO500 bioinformatic workflow leverages unique molecular identifier-based error correction and a statistical model of error rates to suppress technical noise arising from sequencing and FFPE artifacts. To exclude germline variants, the

TSO500 workflow uses a database in combination with information about proximal variants. These combined strategies allow the TSO500 workflow to achieve high specificity and sensitivity for detecting variants with low allele frequencies in FFPE samples. Exclusion of technical noise and germline variants also enables TSO500 to determine MSI status and TMB scores that are highly concordant with gold standard approaches.

Methods

Sequencing

FFPE samples from a total of 170 patients, and included a range of tumor types: colon, gastric, lung, melanoma, uterine, and endometrial. DNA was extracted using the QiAMP DNA FFPE Tissue Kit (QIAGEN). Manufacturer recommendations were followed for quantitative and qualitative evaluation of DNA quality. Nucleic acids were quantified using a Qubit (Invitrogen, Carlsbad, CA). DNA quality was assessed by a Nanodrop spectrophotometer, with an OD 260/280 value between 1.7 and 2.2 considered acceptable. All samples had at least 40 ng/uL of input DNA.

Library preparation was performed using the hybrid capture-based TruSight Oncology 500 Library Preparation Kit (Illumina, San Diego, CA) following the manufacturer's protocol. DNA was fragmented to 90 to 250 bp, with a target peak of around 130 bp. Samples then underwent end repair and A-tailing. Next, adapters containing UMIs were ligated to the ends of the DNA fragments. After a purification step, the DNA fragments were amplified using primers to add index sequences for sample multiplexing (required for cluster generation). For TSO500 samples, two hybridization/capture steps were performed, and only one hybridization and capture step was performed for WES samples. First, either a pool of oligos specific to the 523 genes targeted by TSO500 (TSO500 samples) or probes from the IDT xGen Exome panel (WES samples) were hybridized to the prepared DNA libraries overnight. Next, streptavidin magnetic beads (SMBs) were used to capture probes hybridized to the targeted regions. The hybridization and capture steps were repeated using the enriched DNA libraries to ensure high specificity for the captured regions. Primers were used to amplify the enriched libraries before purification using sample

purification beads. The enriched libraries were quantified and each library was normalized to ensure a uniform representation in the pooled libraries.

Finally, the libraries were pooled, denatured, and diluted to the appropriate loading concentration. TSO500 libraries were sequenced on a Illumina NextSeq™ 550Dx with a read length of 2x101 bp. Up to 8 TSO500 libraries were sequenced per run. WES libraries were sequenced on an Illumina NovaSeq™ 6000 using 2x101 bp reads using the NovaSeq Xp workflow. Six WES libraries were sequenced per flow cell lane for a total of 24 samples per run (S4 flow cells).

WES alignment and variant calling

Reads were first mapped to the hg19 genome using the Burrows-Wheeler Aligner (26,27). Next, duplicate reads were marked on the basis of the soft-clip adjusted position and orientation of each read. For a set of duplicate reads, the read with the highest total base quality score was selected as the representative read. Using paired tumor and normal samples, somatic and germline variants were called using Strelka2 with the “--exome” option enabled and default values for all other parameters (28). Both germline and somatic variants were then annotated using Nirvana (29).

TSO500 alignment and variant calling

A diagram of the TSO500 bioinformatics pipeline is provided in **Figure 1**; quality control metrics are shown in **Supplemental Table 2**. Data analysis begins with an initial DNA alignment to the hg19 genome using the Burrows-Wheeler Aligner. Next using non-random unique molecular identifiers (NRUMIs), duplicate reads are collapsed to reduce PCR and sequencing errors. In the meantime, read collapsing increases the base quality of real low frequency variations, leading to high sensitivity to call low frequency mutations. The NRUMIs used in TSO500 were designed with a pair-wise edit distance of at least 3, which allows error correction of the UMI itself with up to 2 mismatches/indels. Due to A-tailing during library preparation, there would be a lack of nucleotide heterogeneity at the attachment point if a single length UMI was used, so the method uses variable length UMIs (6 and 7-mers) to increase nucleotide heterogeneity: the 7th base is T for all 6-mers, and for the 7-mers the 7th base is never

T. NRUMIs are designed to have a 3bp edit distance between any pair, and so any 1bp error within a NRUMI sequence can be corrected. Families of duplicate reads are identified on the basis of both their aligned position and having matched NRUMIs, which are added to both ends of DNA molecules during sequencing library preparation. Duplicate read families are then collapsed into a single consensus sequence. Collapsed sequences supported by reads from both the forward and reverse strand are denoted as duplex sequences, whereas sequences supported by reads from just one strand are denoted as simplex sequences.

NRUMI collapsed sequences are then indel realigned and stitched together into a consensus fragment using Gemini (30). Stitching further improves per base accuracy and quality and allows for a more accurate calculation of coverage in the overlapping region between read pairs. Reads ending near detected indels are realigned to remove alignment artifacts. Somatic variant calling is then performed using Pisces (31) to identify candidate variants. Variants that do not overlap a region targeted by the TSO500 assay are removed. Post-processing of the small variant calls removes variants that have error rates below quality thresholds (described in the following section). The final set of variants is then annotated using Nirvana (29).

Technical noise suppression

To achieve high fidelity in variant calls, a base-change specific variant filtering algorithm is applied to remove recurrent artifacts and FFPE deamination artifacts. For each variant of interest, background noise at the same site is estimated from ~60 normal FFPE samples profiled by TSO500 assay. Using a binomial distribution to model the background noise level, we then test the significance of each observed mutant allele depth given the total depth and convert the p-value to a variant quality score (AQ) as $-10 \cdot \log_{10}(p\text{-value})$. In addition, the error rate of each nucleotide change is estimated for each support category – duplex sequences (sequences supported by both the forward and reverse strand), simplex sequences originating from forward strand, and simplex sequences originating from reverse strand - in each sample by using all the positions with an allele frequency less than 5%. **Supplemental Figure 1** shows that the error rate can vary considerably depending on support category and nucleotide change. The error rate is calculated as the total mutant reads divided by total depth. For variants of interest, a likelihood ratio score (LQ) is calculated as follows:

$$\text{Likelihood ratio (LQ)} = -10 \cdot \log_{10} \left(\frac{\text{Likelihood (observed variant is an error} \mid DP, \text{error rate})}{\text{Likelihood (observed variant is a mutation} \mid DP, VAF)} \right)$$

For variants with a Catalogue of Somatic Mutations in Cancer (COSMIC) (32) count > 50, the LQ and AQ thresholds are 20 and the remaining sites are 60.

Germline variant filtering

To calculate TMB, germline variants are removed through a strategy that leverages information from public databases combined with the coverage and allele frequency of variants in the same region. The database filter uses information about observed variation from multiple population groups to filter common germline variants. Variants that have an observed allele count of N (where the pipeline default is 10) or more in either the 1000 genome, gnomAD exome, or gnomAD genome database, are filtered out (**Supplemental Figure 2**) (33,34).

Database strategies alone are not sufficient for filtering out private germline mutations in each individual. For variants remaining after database filtering, we apply an additional filter, which uses the allele frequency of proximal known germline variants to determine whether a variant is likely a germline variant. For a given variant, we identify variants on the same chromosome with an allele frequency within the greater of 2 standard deviations (assuming a binomial distribution using coverage and allele frequency of the given variant) and 0.05. If there are more than 5 variants total and more than 95% of them are found in the germline database, then the variant of interest is considered to be likely germline. Additionally, variants with an allele frequency $\geq 90\%$ are labeled as likely germline. In **Supplemental Figure 2**, we show the proximal variants (circled variants) that are within allele frequency 0.05 (black bars) for a germline variant not filtered using database allele counts. This is a simple yet effective approach that accounts for both unfiltered variants in expected germline frequency ranges and germline variants with allele frequency shifts due to copy number variant (CNV) events. One caveat of proximity filter is that it may filter out real somatic mutations when the tumor purity is extremely high ($\geq 80\%$), leading to under-estimation of tumor mutational burden (TMB). We also note that germline filtering is for the purpose of TMB estimation. The classification of germline versus somatic is not perfect, and it is not meant for identification of somatic mutations per se.

Tumor mutational burden

TMB metrics are calculated following small variant calling. TMB is the ratio of the number of eligible variants (mutations) to eligible DNA regions (Mb). Eligible variants (numerator) include only coding variants with a frequency of $\geq 5\%$ and coverage ≥ 50 reads. Single-nucleotide variants (SNVs) and indels are included, but multi-nucleotide variants (MNVs) and variants with a COSMIC count of ≥ 50 are excluded. Variants in blacklisted regions with poor mappability are also excluded. For the denominator, all eligible coding regions (with coverage $\geq 50\times$) are included, except for the blacklist regions. The TSO500 bioinformatics workflow outputs both the TMB score and details of the variants that are included in the TMB calculation.

Microsatellite instability status.

TSO500 determines microsatellite instability status from stitched reads from a tumor sample only. Unstable microsatellite sites are detected by assessing the shift in the length of a microsatellite site for a tumor sample against a set of normal baseline samples. First, the baseline Jensen-Shannon distance, d , is calculated for each pair of baseline samples i, j yielding $D_{i,j}$. Next, Jensen-Shannon distance is calculated between the tumor sample x and each baseline sample, yielding $D_{x,i}$. These two distributions are then compared using a one-sided t-test under the null hypothesis $D_{i,j} \geq D_{x,i}$. Sites with $p < 0.01$ and $D_{x,i} - D_{i,j} > 0.01$ are reported as unstable. We require that each MSI site assessed have at least 60 full-spanning reads. The proportion of unstable MSI sites to total assessed MSI sites is reported as a sample-level microsatellite score, $\text{MSI score} = \frac{\text{Number of unstable MSI sites}}{\text{Total assessed MSI sites}}$. We require at least 40 sites to be assessed to determine a MSI score.

To ensure that microsatellite instability can be assessed in a wide range of tumor types, the TSO500 assay initially targeted 175 noncoding homopolymer regions. We observed that certain homopolymer regions tended to test as unstable more frequently for samples with a particular ethnic background. Using all sites, we observe that African samples tended to have considerably higher MSI scores in comparison to samples from other ethnic backgrounds (**Supplemental Figure 4**, n=101 European, n=23 East Asian, n=10 African, and m=6 Hispanic). After exclusion of sites that showed an ethnicity bias, we observe more comparable MSI scores for samples with

different ethnic backgrounds. To ensure robust MSI status calculation, we also exclude sites with low sequencing coverage. Using these two criteria for site selection, we assess a total of 130 noncoding homopolymer regions to determine MSI status.

Results

Technical sequencing noise suppression

High accuracy biomarker calculation is dependent on high-accurate sequencing, alignment and variant calling. The TSO500 assay integrates UMIs for error correction, allowing for suppression of PCR duplicates and sequencing errors, while preserving DNA molecules supporting low frequency variants. **Figure 2** presents consolidated error rates before and after UMI read collapsing for 170 normal FFPE samples. We observe a systematic reduction in error rates across all samples that is independent of overall sample error rate. The mean error rate for all bases in the panel was 0.074% before condensing and 0.032% after condensing. UMI based error correction performance improved with greater family depth (**Supplemental Figure 3**). Mean family depth was 4.788 across all samples.

UMI based read collapsing suppresses sequencing error, but there are additional sources of false positive calls that are quality-dependent and/or caused by DNA damage. For example, deamination (C:G→T:A) and oxidation (C:G→A:T) events were observed with varying rates depending on the strand and whether a read was simplex or duplex (**Supplemental Figure 1**). Therefore, we use normal FFPE samples to estimate the background noise at each base and model the confidence of each variant call. We also calculate the error rate of each nucleotide change for each sample to estimate a sample-specific error profile (see methods). Using the sample-specific error profile, we calculate a sample-specific variant calling threshold, allowing us to achieve comparable variant calling performance between samples of varying quality. By comparing the likelihood that a variant occurs at a given position against the likelihood that an error has occurred at a given position, we can further eliminate low quality, false positive variant calls. **Figure 3** gives the distribution of total false positive variant calls for 170 normal FFPE samples across the entire 1.94 Mb TSO500 panel; overall, we achieve specificity > 99.9998% in these samples.

To estimate our limit of detection and sensitivity for small variant calling, we performed a series of dilution experiments using 10 FFPE tissue samples and 4 commercial reference standards. Variant allele frequencies (VAFs) for 35 variants were determined in undiluted samples using digital droplet PCR and ranged from 21-78%. These samples were then diluted so that VAFs ranged as low as 1-2% and that the majority of variants had $VAF \leq 6\%$. Summary statistics for our variant calling performance in these diluted samples are shown in **Table 1**. For variants with VAF of $5 \pm 2.5\%$ we achieve sensitivity of 99%.

Germline variant filtering

As TSO500 is a tumor only workflow, our variant calling algorithms cannot exclude germline variants by using information from a matched normal sample. Instead, identification of germline variants is accomplished by using information from public databases of germline variants (33,34) and a second line proximity filter. By excluding variants that appear more than 10 times in public databases, we were able to identify, on average, 99.73% of the coding region variants as germline variants in each sample; germline variants were verified by assaying the matched normal samples using with TSO500. (33,34). While database filtering is effective, a median of 4 germline variants, not represented in databases, remained unfiltered in our set of 170 tumor samples. And so, we applied an additional filter that uses the database information of proximal variants (see methods). After applying the proximity filter, the number of remaining germline variants in each sample, germline residuals, was further reduced to a median of 1 per sample (**Figure 4**). We calculated and compared TMB scores for each sample using the tumor sample only and using both the tumor and normal sample. The high correlation of the tumor-only TMB values with the matched tumor-normal values (adjusted $R^2 = 0.9945$) further demonstrates the effectiveness of our germline filters and highlights that our tumor only workflow can accurately estimate TMB (**Figure 5A**).

Tumor mutational burden using tumor sample only

Given the formulation of TMB, high quality alignment and variant calling with TSO500 should enable robust TMB calculation. The net effect of UMI based read collapsing and technical noise and germline variant filtering is a reduction in false positive variants in a typical FFPE sample

from approximately 1500/Mb to < 5/Mb. To assess whether this reduction in false positives translated into reliable TMB values concordant with WES sequencing, we re-processed 108 out of 170 of our samples with whole exome sequencing (matched tumor-normal samples) and calculated the nonsynonymous TMB, which is the typically reported TMB value for WES data (35-37). We then compared the WES tumor-normal TMB values with TSO500 tumor-only TMB values (by default TSO500 uses both coding SNVs and Indels regardless of functional impact, total TMB), observing a high correlation between the two sets of TMB values (adjusted $R^2=0.933$, **Figure 5B**). We observe that TSO500 TMB values tend to be elevated relative to WES TMB (intercept=3.201), which is likely attributable to the use of total TMB instead of only nonsynonymous variants for WES. Similarly, the elevated slope of the best fit line (1.36) is attributable to the same reason.

We next assessed the reproducibility of TSO500 tumor-only TMB scores. Four FFPE samples and four commercially available control samples were selected to be assayed across a variety of conditions. Using each of these eight FFPE samples, three operators generated libraries using three lots of reagents, three sets of instruments with each combination of reagents in duplicate, totaling to 36 technical replicates per sample. Across all eight samples, we observed highly similar values across all operators and conditions (**Supplemental Figure 6**); the mean coefficient of variation across all samples was 0.0567.

Microsatellite instability status using tumor sample only

The TSO500 bioinformatics workflow leverages UMI error-corrected stitched reads to determine MSI status from a tumor sample only. To ensure that TSO500 can determine MSI status in a wide range of tumor sites, a total of 130 noncoding homopolymer regions are assessed for microsatellite instability. We also specifically exclude sites demonstrating an ethnicity bias (**Supplemental Figure 4**). To enable a tumor-only determination of MSI status, TSO500 uses an information theory-based approach that compares the distribution of site differences between the tumor sample and a panel of normal samples with the distribution of differences between samples within the panel of normal (see methods). From our initial set of 170 FFPE samples, we assayed 106 samples for MSI status using an MSI-PCR assay commercially available from Promega that assess five mononucleotide sites (**Figure 6**). Using a MSI score cutoff to separate

MSI-high from MSI-stable samples, the TSO500 MSI score showed 98% overall agreement with MSI-PCR (**Supplemental Table 3**).

To examine the reproducibility of determining MSI status using the tumor only workflow of TSO500, we performed a total of 36 technical replicate experiments for eight samples (four FFPE samples and four commercially available controls). Each sample was processed by three operators who used three lots of reagents and three sets of instruments to perform each experiment twice. We observed robust MSI scores across all operators and conditions (**Supplemental Figure 7**) and the mean standard deviation across all samples was 0.0196.

Discussion

The treatment of cancer is moving towards the assignment of treatment based on the genetic attributes of an individual patient's tumor. Given the wide landscape of cancerous genetic aberrations and treatments specifically targeting a specific mechanism, there is a pressing need for cost effective methods to systematically characterize a tumor. Here we present the bioinformatics solutions of TSO500, a cost-effective tumor only pan-cancer NGS assay, which enables the identification of a range of DNA variant types (including SNVs, MNVs, indels, and complex indels), as well as the determination of MSI and TMB. We describe the bioinformatic algorithms underlying our error correction, variant calling, and determination of biomarkers using a tumor sample only. In addition to showing that tumor only analysis with TSO500 enables accurate tumor variant calls, we also demonstrate that biomarker metrics reported by TSO500 are highly concordant with gold standard methods and are highly reproducible. Our bioinformatics solutions are available as Linux containers (Docker and Singularity images, support.illumina.com/sequencing/sequencing_kits/trusight-oncology-500/questions.html), which enables on-premise software deployment when security is a high priority or when bioinformatics expertise is not readily available (38).

The tumor-only workflow of TSO500 is designed to be effective for research applications today, as well as for the development of future applications. TSO500 enables research into emerging biomarkers such as MSI and TMB status. TMB has been demonstrated to be useful for determining the success of checkpoint inhibitor immunotherapies in NSCLC patients (20-22).

Similarly, there is evidence that MSI status may identify CRC patients that respond better to checkpoint inhibitor immunotherapies (14,15,39). Despite the promise of these biomarkers, we note that the National Comprehensive Cancer Network (NCCN) guidelines for using these biomarkers are still emerging, partly because methods for determining these biomarkers are still emerging (40). Here, we fully describe the algorithms underlying our TMB and MSI calculations to enable further research and development of these biomarkers. Furthermore, data from the TSO500 workflow represents a systematic interrogation of the mutational landscape of a tumor and can be leveraged to develop future biomarkers. We are in the process of developing an in vitro diagnostic (IVD) based on TSO500 content. This IVD is anticipated to have the same technical capabilities described here and will enable clinical application of this technology.

We envision several additional applications using TSO500. Several variant types not described in detail here – including RNA variants, copy number variation, and splice variants - can also be detected using TSO500. These additional variant types may contribute to future biomarkers as research progresses (41-44). We are also developing a circulating tumor DNA (ctDNA) version of TSO500 because there is increasing evidence that supports the use of liquid biopsies, and specifically the analysis of ctDNA, to diagnose and monitor the treatment of cancer patients (45-50). Next-generation sequencing enables characterization of each patient's tumor to a level not previously achievable, and we expect that these technologies will accelerate progress towards the implementation of precision medicine in oncology.

Conclusion

TSO500 is an accurate tumor-only workflow that enables researchers to systematically characterize tumors and identify the next generation of clinical biomarkers. We envision the broader adoption of comprehensive genomic profiling panels like TSO500 will further help advance precision medicine.

List of abbreviations

microsatellite instability (MSI); tumor mutational burden (TMB); TruSight™ Oncology 500 (TSO500); formalin-fixed paraffin-embedded (FFPE); unique molecular identifier (UMI); next-generation sequencing (NGS); immunohistochemistry (IHC); polymerase chain reaction (PCR); whole exome sequencing (WES); non-random unique molecular identifiers (NRUMIs); variant

quality score (AQ); likelihood ratio score (LQ); Catalogue of Somatic Mutations in Cancer (COSMIC); copy number variant (CNV); Variant allele frequencies (VAFs)

Declarations

Ethics approval and consent to participate

The FFPE samples used in this study were purchased from commercial vendors.

Consent to publish

Not applicable

Availability of data and materials

The dataset will be available after being authorized by the corresponding author on reasonable request.

Competing interests

All of the authors are employees of and own stock in Illumina, Inc.

Funding

This research was supported by Illumina, Inc.

Authors' Contributions

Study concept and design: CZ, JL, JD, TP, SB. Algorithm development and data analysis: CZ, TJ, JHJ, SZ, JT, YF. The authors have read and approved the final manuscript.

Acknowledgements

None

References

1. Stegmeier F, Warmuth M, Sellers WR, Dorsch M. Targeted cancer therapies in the twenty-first century: lessons from imatinib. *Clinical pharmacology and therapeutics* **2010**;87:543-52.
2. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **2006**;439:358-62.
3. Ji H, Li D, Chen L, Shimamura T, Kobayashi S, McNamara K, *et al.* The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies. *Cancer cell* **2006**;9:485-95.
4. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **2007**;448:561-6.
5. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clinical chemistry* **2015**;61:64-71.
6. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *The Journal of molecular diagnostics : JMD* **2013**;15:623-33.
7. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **2018**;173:355-70.e14.
8. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, *et al.* Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell systems* **2018**;6:271-81.e7.
9. Payne K, Gavan SP, Wright SJ, Thompson AJ. Cost-effectiveness analyses of genetic and genomic diagnostic tests. *Nature Reviews Genetics* **2018**;19:235.

10. Administration USFaD. 2017 June 7, 2019. FDA grants accelerated approval to pembrolizumab for first tissue/site agnostic indication.
<<https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-pembrolizumab-first-tissuesite-agnostic-indication>>. Accessed 2019 June 7, 2019.
11. Diaz LA, Jr., Le DT. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine* **2015**;373:1979.
12. Ward R, Meagher A, Tomlinson I, O'Connor T, Norrie M, Wu R, *et al.* Microsatellite instability and the clinicopathological features of sporadic colorectal cancer. *Gut* **2001**;48:821-9.
13. Boussios S, Ozturk MA, Moschetta M, Karathanasi A, Zakynthinakis-Kyriakou N, Katsanos KH, *et al.* The Developing Story of Predictive Biomarkers in Colorectal Cancer. *Journal of personalized medicine* **2019**;9.
14. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (New York, NY)* **2017**;357:409-13.
15. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine* **2015**;372:2509-20.
16. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer research* **1998**;58:5248-57.
17. McConechy MK, Talhouk A, Li-Chang HH, Leung S, Huntsman DG, Gilks CB, *et al.* Detection of DNA mismatch repair (MMR) deficiencies by immunohistochemistry can

- effectively diagnose the microsatellite instability (MSI) phenotype in endometrial carcinomas. *Gynecologic oncology* **2015**;137:306-10.
18. Hempelmann JA, Lockwood CM, Konnick EQ, Schweizer MT, Antonarakis ES, Lotan TL, *et al.* Microsatellite instability in prostate cancer by PCR or next-generation sequencing. *J Immunother Cancer* **2018**;6:29.
 19. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen H, *et al.* Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precision Oncology* **2017**:1-15.
 20. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, *et al.* Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer cell* **2018**;33:843-52.e4.
 21. Hellmann MD, Rizvi NA, Goldman JW, Gettinger SN, Borghaei H, Brahmer JR, *et al.* Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, phase 1, multicohort study. *The Lancet Oncology* **2017**;18:31-41.
 22. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, *et al.* First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *The New England journal of medicine* **2017**;376:2415-26.
 23. Truesdell J, Miller VA, Fabrizio D. Approach to evaluating tumor mutational burden in routine clinical practice. *Translational lung cancer research* **2018**;7:678-81.
 24. Allgauer M, Budczies J, Christopoulos P, Endris V, Lier A, Rempel E, *et al.* Implementing tumor mutational burden (TMB) analysis in routine diagnostics-a primer for molecular pathologists and clinicians. *Translational lung cancer research* **2018**;7:703-15.
 25. Buchhalter I, Rempel E, Endris V, Allgauer M, Neumann O, Volckmar AL, *et al.* Size matters: Dissecting key parameters for panel-based tumor mutational burden analysis. *International journal of cancer* **2019**;144:848-58.

26. Li H. 2013 September 25. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <<https://arxiv.org/abs/1303.3997>>. Accessed 2019 September 25.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **2009**;25:1754-60.
28. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* **2018**;15:591-4.
29. Stromberg M, Roy R, Lajugie J, Jiang Y, Li H, Margulies E. Nirvana: Clinical Grade Variant Annotator. Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. Boston, Massachusetts, USA: ACM; 2017. p 596-.
30. Illumina. September 25. Gemini 5.2.10 Design Document. <<https://github.com/Illumina/Pisces/wiki/Gemini-5.2.10-Design-Document>>. Accessed 2019 September 25.
31. Illumina. September 25. Pisces. <<https://github.com/Illumina/Pisces>>. Accessed 2019 September 25.
32. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research* **2018**;47:D941-D7.
33. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* **2019**:531210.
34. The Genomes Project C, Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, *et al.* A global reference for human genetic variation. *Nature* **2015**;526:68.

35. Campesato LF, Barroso-Sousa R, Jimenez L, Correa BR, Sabbaga J, Hoff PM, *et al.* Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. *Oncotarget* **2015**;6.
36. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *The New England journal of medicine* **2018**;378:2093-104.
37. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome medicine* **2017**;9:34.
38. Matelsky J, Kiar G, Johnson E, Rivera C, Toma M, Gray-Roncal W. Container-Based Clinical Solutions for Portable and Reproducible Image Analysis. *Journal of digital imaging* **2018**;31:315-20.
39. National Comprehensive Cancer Network. 2019 June 7, 2019. NCCN Clinical Practice Guidelines in Oncology: Colon Cancer v2.2019. https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf. Accessed 2019 June 7, 2019.
40. Network NCC. 2019 June 7, 2019. NCCN Clinical Practice Guidelines in Oncology: Non-Small Cell Lung Cancer. https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf. Accessed 2019 June 7, 2019.
41. Horvath A, Pakala SB, Mudvari P, Reddy SDN, Ohshiro K, Casimiro S, *et al.* Novel Insights into Breast Cancer Genetic Variance through RNA Sequencing. *Scientific reports* **2013**;3:2256.
42. Watanabe T, Ueno H, Watabe Y, Hiraoka N, Morizane C, Itami J, *et al.* ACTN4 copy number increase as a predictive biomarker for chemoradiotherapy of locally advanced pancreatic cancer. *British journal of cancer* **2015**;112:704-13.

43. Stronach EA, Paul J, Timms KM, Hughes E, Brown K, Neff C, *et al.* Biomarker Assessment of HR Deficiency, Tumor *BRCA1/2* Mutations, and *CCNE1* Copy Number in Ovarian Cancer: Associations with Clinical Outcome Following Platinum Monotherapy. *Molecular Cancer Research* **2018**;16:1103-11.
44. Licitra L, Mesia R, Rivera F, Remenár É, Hitt R, Erfán J, *et al.* Evaluation of EGFR gene copy number as a predictive biomarker for the efficacy of cetuximab in combination with chemotherapy in the first-line treatment of recurrent and/or metastatic squamous cell carcinoma of the head and neck: EXTREME study. *Annals of Oncology* **2010**;22:1078-87.
45. Lanman RB, Mortimer SA, Zill OA, Sebisanoovic D, Lopez R, Blau S, *et al.* Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA. *PloS one* **2015**;10:e0140712.
46. Earl J, Garcia-Nieto S, Martinez-Avila JC, Montans J, Sanjuanbenito A, Rodríguez-Garrote M, *et al.* Circulating tumor cells (CTC) and KRAS mutant circulating free DNA (cfDNA) detection in peripheral blood as biomarkers in patients diagnosed with exocrine pancreatic cancer. *BMC Cancer* **2015**;15:797.
47. Fernandez-Cuesta L, Perdomo S, Avogbe PH, Leblay N, Delhomme TM, Gaborieau V, *et al.* Identification of Circulating Tumor DNA for the Early Detection of Small-cell Lung Cancer. *EBioMedicine* **2016**;10:117-23.
48. Babayan A, Pantel K. Advances in liquid biopsy approaches for early detection and monitoring of cancer. *Genome medicine* **2018**;10:21.
49. Jamal-Hanjani M, Wilson GA, Horswell S, Mitter R, Sakarya O, Constantin T, *et al.* Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. *Annals of Oncology* **2016**;27:862-7.
50. Pantel K. Blood-Based Analysis of Circulating Cell-Free DNA and Tumor Cells for Early Cancer Detection. *PLOS Medicine* **2016**;13:e1002205.

Table 1. Summary statistics for limit of detection for small DNA variants. Abbreviations: MNV = multi-nucleotide variant, SNV = single-nucleotide variant, VAF = variant allele frequency.

Expected VAF*	Variant Type	Called Samples / Total Samples	Sensitivity (%)
10.0% \pm 2.5%	SNV	18/18	100.0
	Insertion	4/4	100.0
	Deletion	4/4	100.0
	Complex Indel	10/10	100.0
	All	36/36	100.0
5.0% \pm 2.5%	SNV	433/434	99.8
	MNV	4/4	100.0
	Insertion	25/26	96.2
	Deletion	104/108	96.3
	Complex Indel	10/10	100.0
	All	576/582	99.0

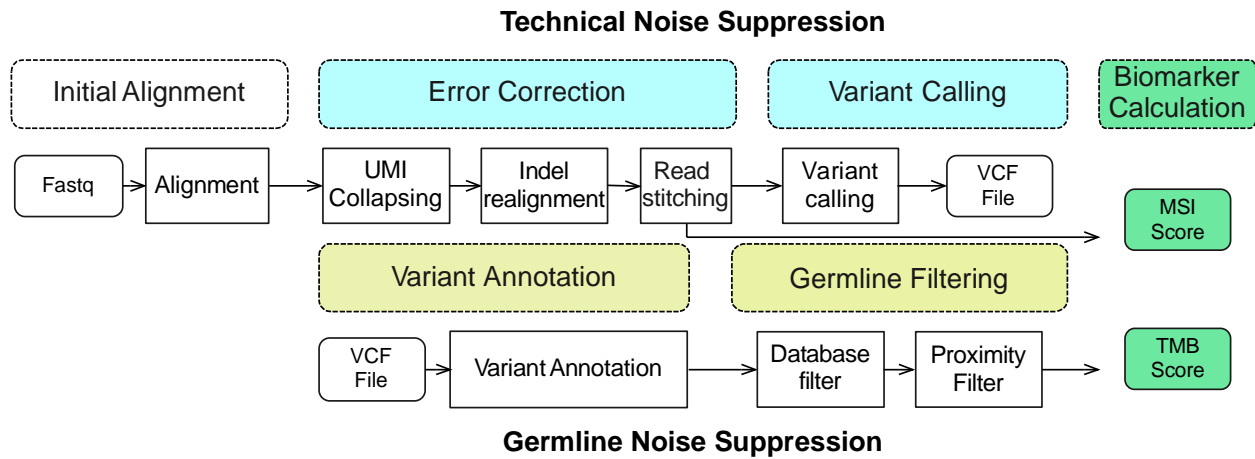


Figure 1. TSO500 bioinformatic workflow for variant calling and biomarker calculation .
Abbreviations: Indel = insertion/deletion, MSI = microsatellite instability, TMB = tumor mutational burden, UMI = unique molecular identifier.

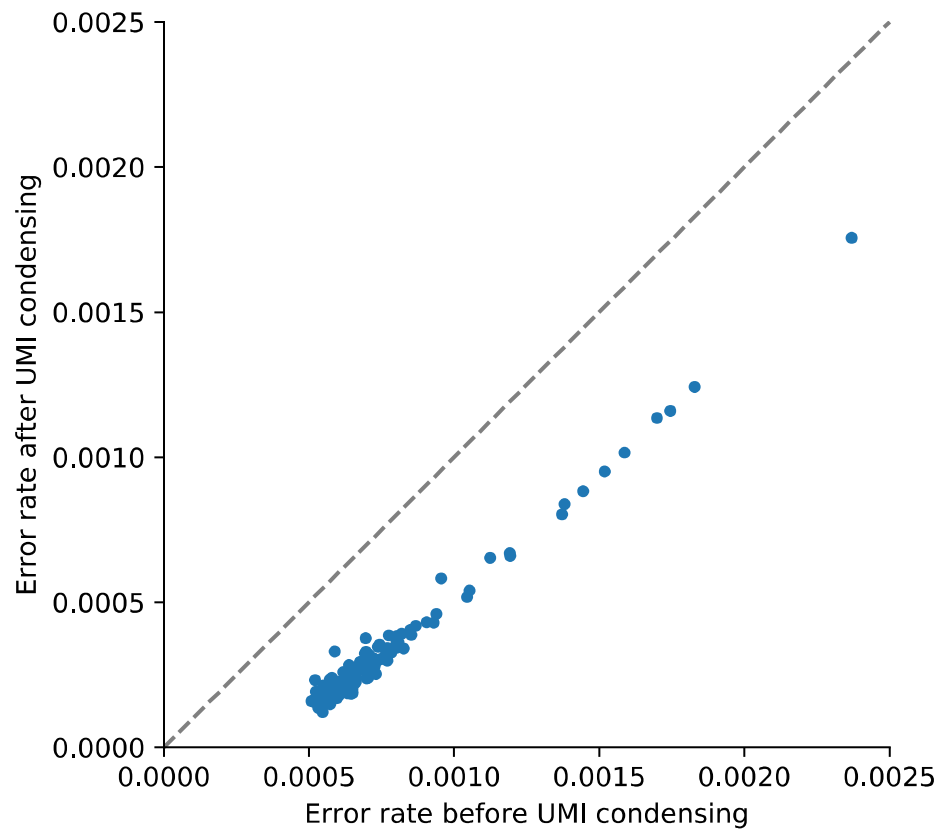


Figure 2. Reduction in error rates using unique molecular identifiers. Sequencing error rates for 170 normal FFPE samples before and after condensing reads using unique molecular identifiers.

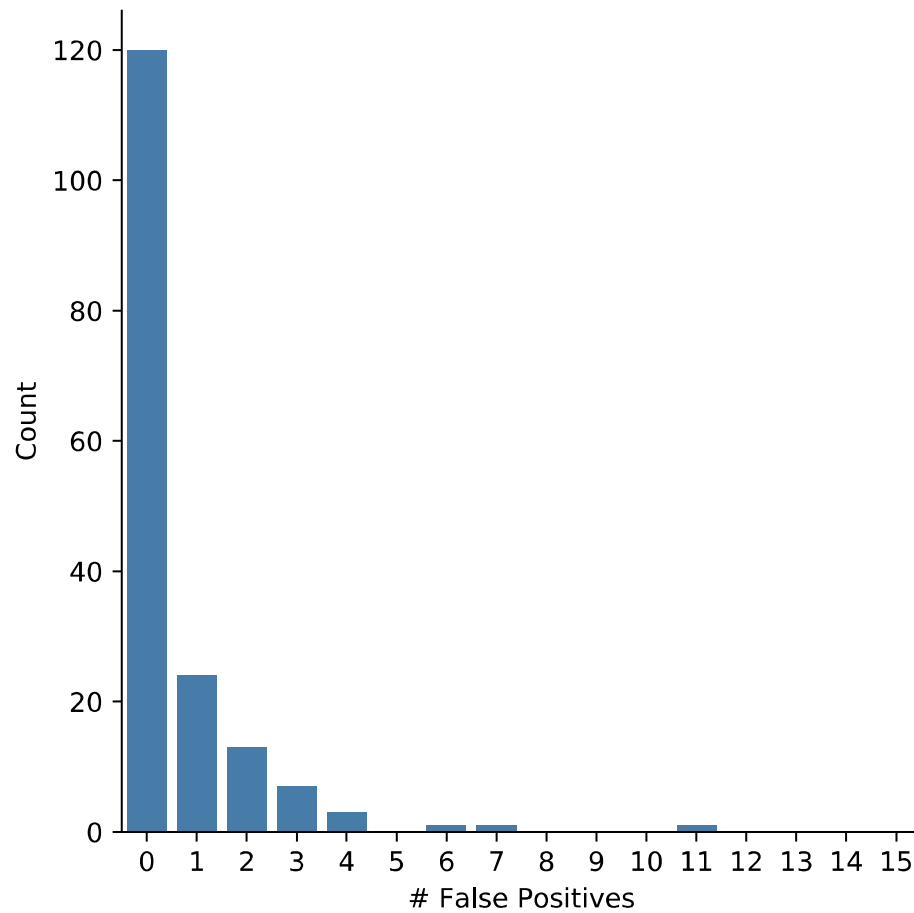


Figure 3. Distribution of the number of putative false positive variant calls in 170 normal FFPE tissue samples. The vertical axis gives the number of samples.

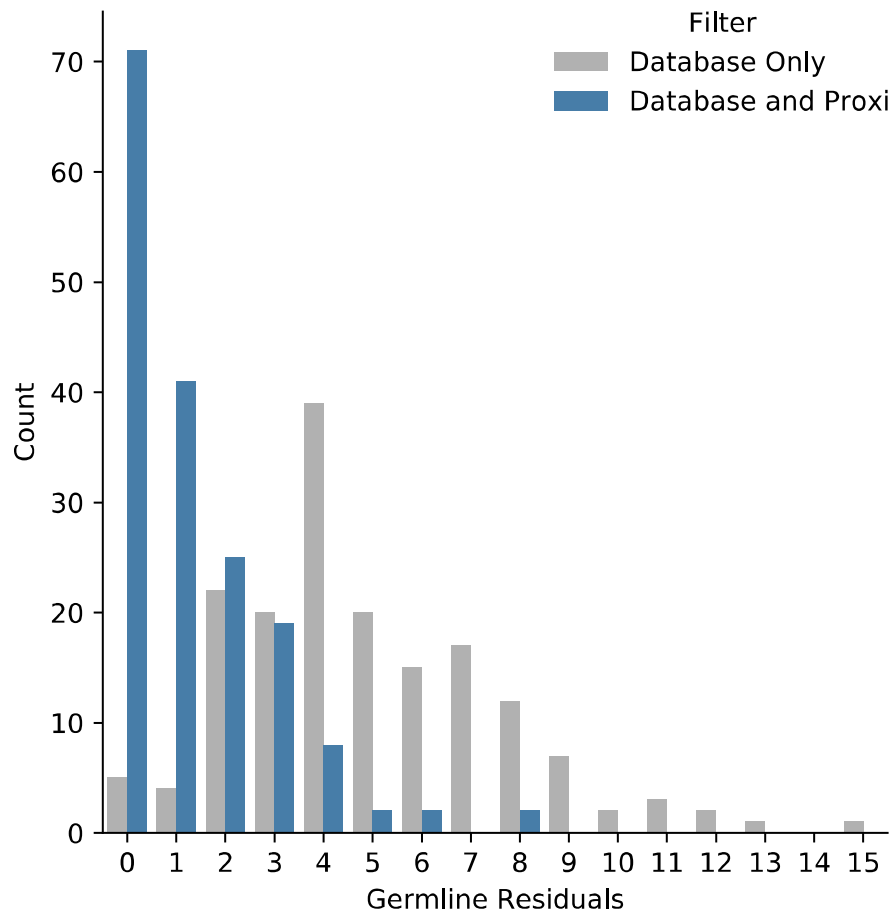


Figure 4. Reduction in putative germline variants using the proximity filter for 170 tumor FFPE tissue samples. Germline variants were determined by comparing to a matched normal FFPE tissue sample. The horizontal axis gives the number germline variants remaining either after database filtering only or after database and proximity filtering. The vertical axis gives the number of samples.

TruSight Oncology 500

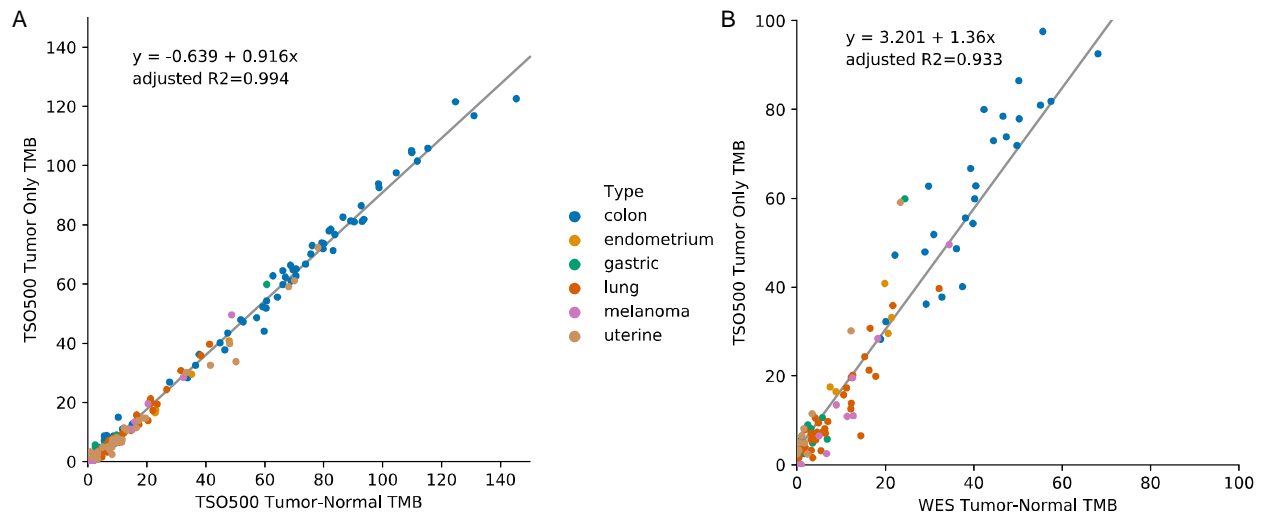


Figure 5. Comparison of TSO500 tumor-only TMB values to tumor-normal TMB values.

A. TMB values using the TSO500 workflow are shown for 170 Tumor-Normal pairs. Tumor-only TMB is strongly correlated with Tumor-Normal TMB in formalin-fixed, paraffin-embedded tissue samples (adjusted $r^2 = 0.9945$). B. TMB values are shown for 108 Tumor-Normal pairs. Tumor-only TSO500 TMB is strongly correlated with WES TMB in formalin-fixed, paraffin-embedded tissue samples (adjusted $r^2 = 0.933$). TSO500 TMB is calculated using both SNVs and Indels. WES TMB is calculated using nonsynonymous variants only. Abbreviations: TMB = tumor mutational burden, WES = whole exome sequencing.

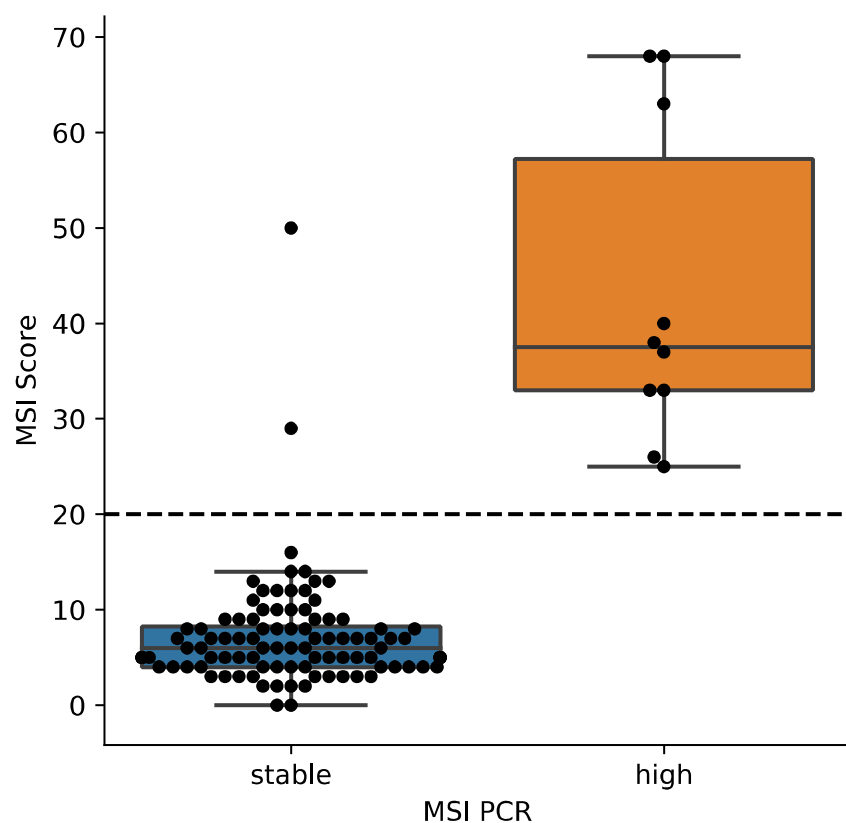


Figure 6. Comparison of tumor-normal MSI-PCR and tumor-only TSO500 MSI-score. MSI status determined by Promega MSI-PCR using matched tumor and normal samples is shown on the horizontal axis for 106 FFPE samples from a variety of tumor types (colon, gastric, lung, melanoma, uterine, endometrium). The vertical axis gives the TSO500 tumor-only MSI score. The dashed line indicates the MSI score cutoff for determining MSI-high samples.

Supplemental Table 1. TSO500 assay content (523 genes).

Small DNA Variants

ABL1	CSF3R	FLI1	GATA4	IRS1	NTRK1
ABL2	CSNK1A1	FLT4	GATA6	IRS2	NTRK2
ACVR1	CTCF	FOXA1	GID4	JAK1	NTRK3
ACVR1B	CTLA4	FOXO1	GLI1	JUN	NUP93
ALOX12B	CTNNA1	FOXP1	GNA13	KAT6A	NUTM1
ANKRD11	CUL3	FRS2	GPR124	KDM5A	PAK1
ANKRD26	CUX1	FUBP1	GPS2	KDM5C	PAK3
ARAF	CXCR4	FYN	GREM1	KDM6A	PAK7
ARFRP1	CYLD	GABRA6	GRIN2A	KEAP1	PARK2
ARID1B	DAXX	GATA1	GRM3	KEL	PARP1
ARID2	DCUN1D1	GATA2	GSK3B	KIF5B	PAX3
ARID5B	DDX41	GATA3	H3F3A	KLF4	PAX5
ASXL1	DHX15	GATA4	H3F3B	KLHL6	PAX7
ASXL2	DICER1	GATA6	H3F3C	KMT2B	PAX8
ATRX	DIS3	GID4	HGF	KMT2C	PBRM1
AURKA	DNAJB1	GLI1	HIST1H1C	KMT2D	PDCD1
AURKB	DNMT1	GNA13	HIST1H2BD	LAMP1	PDCD1LG2
AXIN1	DNMT3B	EPHA7	HIST1H3A	LATS1	PDK1
AXIN2	DOT1L	EPHB1	HIST1H3B	LATS2	PDPK1
AXL	E2F3	ERCC3	HIST1H3C	LMO1	PGR
B2M	EED	ERCC4	HIST1H3D	LRP1B	PHF6
BBC3	EGFL7	ERCC5	HIST1H3E	LYN	PHOX2B
BCL10	EIF1AX	ERRFI1	HIST1H3F	LZTR1	PIK3C2B
BCL2L1	EIF4A2	ETS1	HIST1H3G	MAGI2	PIK3C2G
BCL2L11	EIF4E	ETV1	HIST1H3H	MALT1	PIK3C3
BCL2L2	EML4	ETV4	HIST1H3I	MAP2K4	PIK3R2
BCOR	EPCAM	ETV5	HIST1H3J	MAP3K1	PIK3R3
BCORL1	EPHA3	ETV6	HIST2H3A	MAP3K13	PIM1
BCR	EPHA5	EWSR1	HIST2H3C	MAP3K14	PLCG2
BIRC3	EPHA7	FAM123B	HIST2H3D	MAP3K4	PLK2
BLM	EPHB1	FAM46C	HIST3H3	MAPK1	PMAIP1
BMPR1A	ERCC3	FANCA	HLA-A	MAPK3	PMS1
BRD4	ERCC4	FANCC	HLA-B	MAX	PNRC1
BTG1	ERCC5	FANCD2	HLA-C	MDC1	POLD1
C11orf30	ERRFI1	FANCE	HNRNPK	MED12	POLE
CALR	ETS1	FANCF	HOXB13	MEF2B	PPARG
CASP8	ETV1	FANCG	HSD3B1	MEN1	PPM1D
CBFB	ETV4	FAS	HSP90AA1	MGA	PPP2R1A
CBL	ETV5	FAT1	ICOSLG	MITF	PPP6C
CD274	ETV6	FGF19	ID3	MST1	PRDM1
CD276	EWSR1	FH	IFNGR1	MST1R	PREX2
CD74	FAM123B	FLCN	IGF1	MYB	PRKAR1A
CDC73	FAM46C	FLI1	IGF1R	MYOD1	PRKCI
CDK8	FANCA	FLT4	IGF2	NAB2	PRKDC
CDKN1A	FANCC	FOXA1	IKBKE	NCOA3	PRSS8
CDKN1B	FANCD2	FOXO1	IKZF1	NCOR1	PTPRD
CDKN2B	FANCE	FOXP1	IL10	NEGR1	PTPRS
CDKN2C	FANCF	FRS2	IL7R	NF2	PTPRT
CENPA	FANCG	FUBP1	INHA	NFE2L2	QKI
CHD2	FAS	FYN	INHBA	NFKBIA	RAB35
CHD4	FAT1	GABRA6	INPP4A	NKX2-1	RAC1
CIC	FGF19	GATA1	INSR	NKX3-1	RAD21
CRKL	FH	GATA2	IRF2	NOTCH4	RAD50
CRLF2	FLCN	GATA3	IRF4	NSD1	RAD52

TruSight Oncology 500

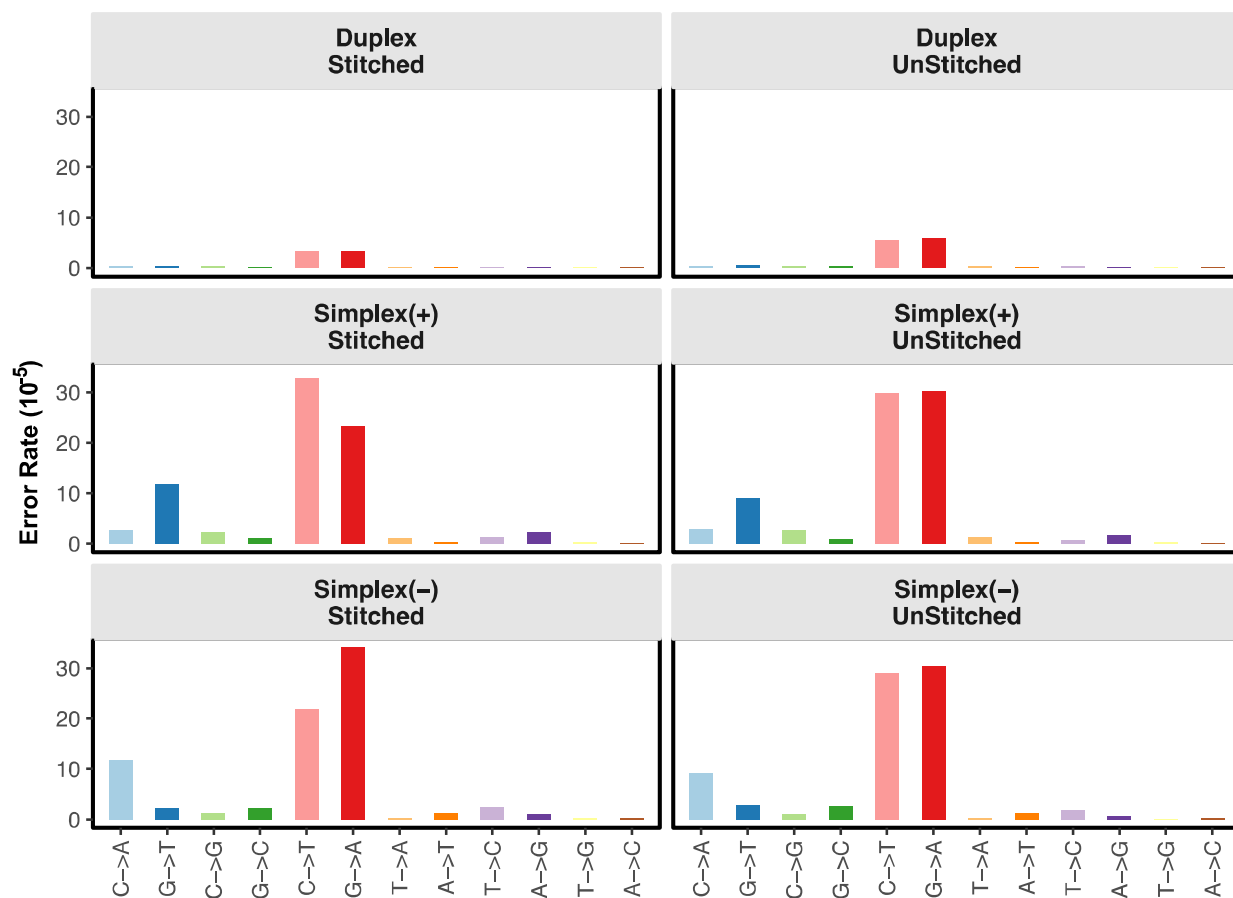
RAF1	RUNX1T1	SMARCA4	STAT3	TFRC	WISP3
RANBP2	RYBP	SMARCD1	STAT4	TGFBR1	WT1
RARA	SDHA	SMC1A	STAT5A	TGFBR2	XIAP
RASA1	SDHAF2	SMC3	STAT5B	TMEM127	XPO1
RBM10	SDHB	SNCAIP	STK40	TMPRSS2	YAP1
RECQL4	SDHC	SOCS1	SUFU	TNFAIP3	YES1
REL	SDHD	SOX10	SUZ12	TNFRSF14	ZBTB2
RFWD2	SETBP1	SOX17	SYK	TOP1	ZBTB7A
RHEB	SETD2	SOX2	TAF1	TOP2A	ZFXH3
RHOA	SF3B1	SOX9	TBX3	TP63	ZNF217
RIT1	SH2B3	SPEN	TCEB1	TRAF2	ZNF703
RNF43	SH2D1A	SPOP	TCF3	TRAF7	ZRSR2
RPS6KA4	SHQ1	SPTA1	TCF7L2	TSHR	
RPS6KB2	SLIT2	SRSF2	TERC	U2AF1	
RPTOR	SMAD2	STAG1	TET1	VEGFA	
RUNX1	SMAD3	STAG2	TFE3	VTCN1	

Supplemental Table 2. TSO500 quality control metrics.

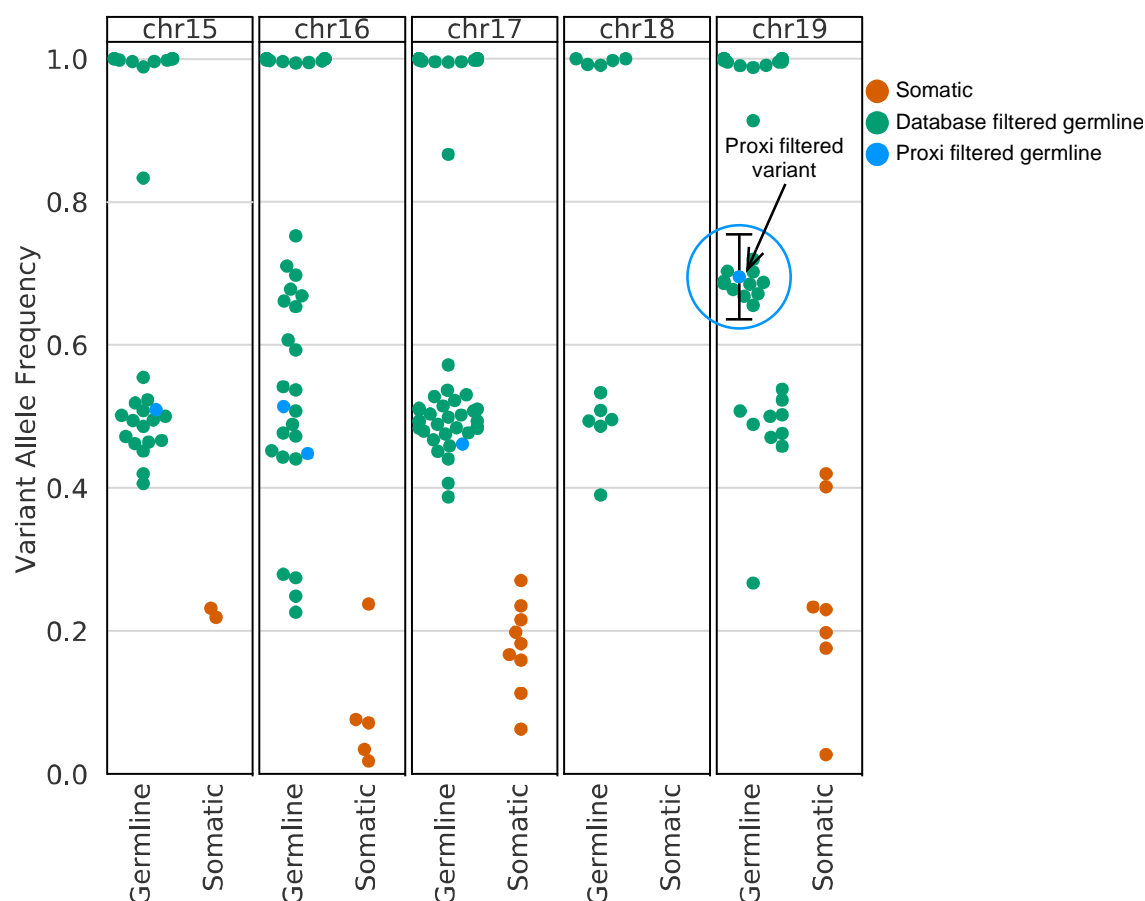
Category	Metrics	Cutoff
Sequencing Metrics	Percent Q30 read1	$\geq 80\%$
	Percent Q30 read2	$\geq 80\%$
	Percent PF reads	$\geq 80\%$
Sample Metrics	Median Insert Size	≥ 70
	Median Exon Coverage	$\geq 150x$
	Percent Exon 50x	$\geq 90\%$
	Usable MSI Sites	≥ 40
	Contamination	Contamination Score ≤ 3106 or re-arrangement p-value < 0.05

Abbreviation: MSI = microsatellite instability.

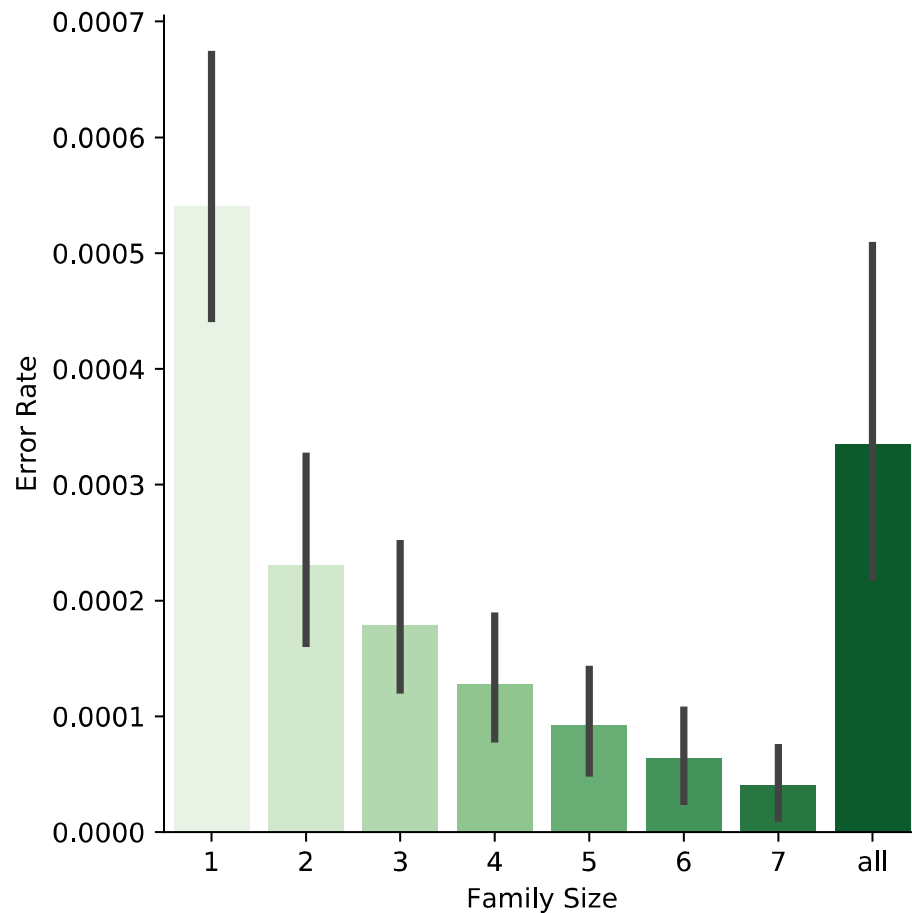
Supplemental Table 3. Details of MSI samples.



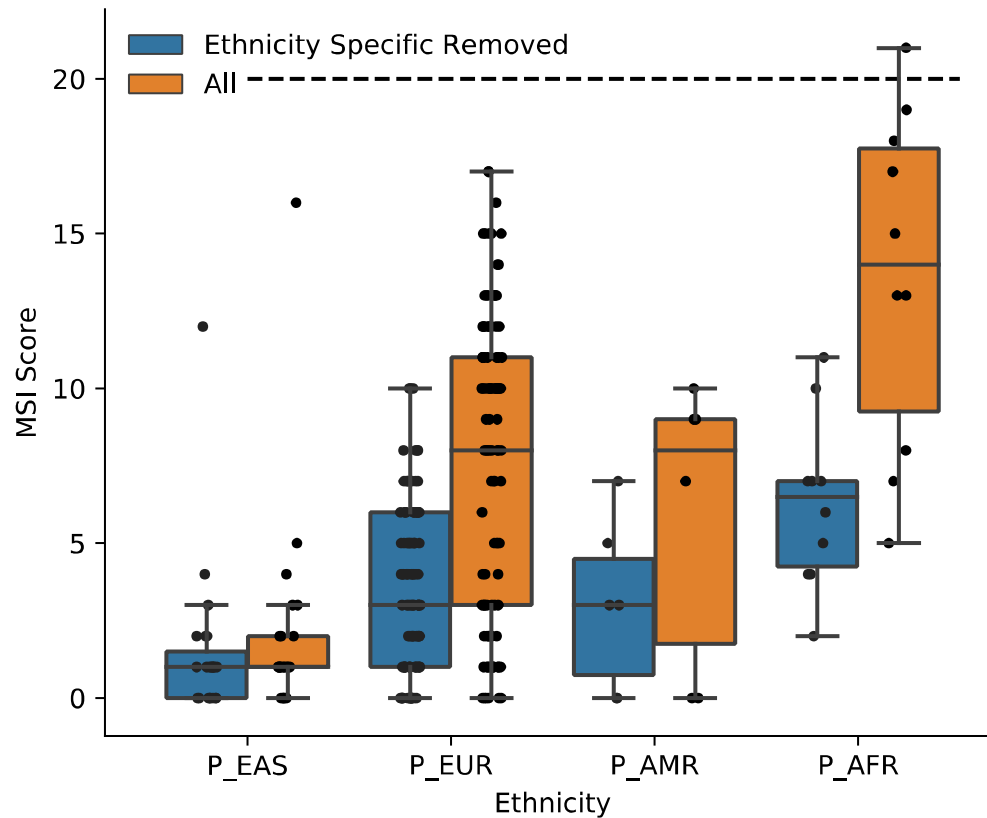
Supplemental Figure 1. Error rates from the simplex forward, simplex reverse, and duplex strands with and without stitching. Deamination errors (G→A), which are prevalent in formalin-fixed, paraffin-embedded tissue and oxidation errors resulting from with sonication (C→A or G→T errors) are greatly reduced in the duplex, stitched panel.



Supplemental Figure 2. Germline variant filtering. Variants that are observed 10 times or more in public germline variant databases are filtered. Additional variants are categorized as germline variants using the proximity filter, which uses the database information of variants with a similar allele frequency (circled variants). Black bars indicate allele frequency range of the proximity filter, which is the maximum of 0.05 or 2 standard deviations giving a binomial distribution.

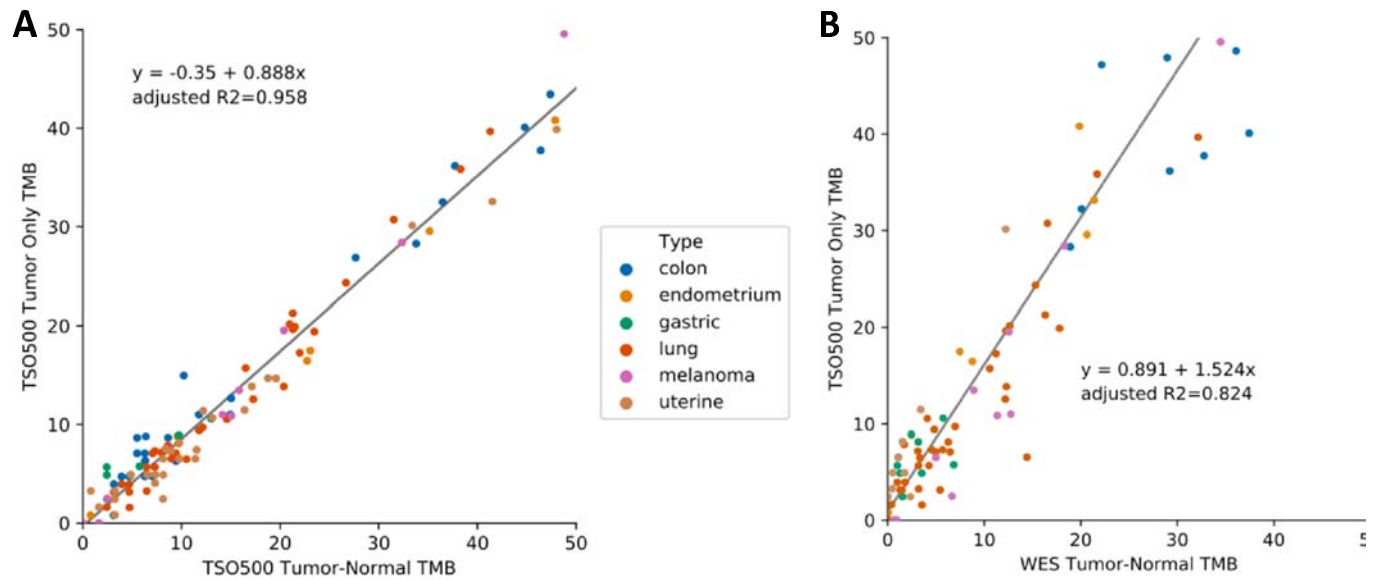


Supplemental Figure 3. Sequencing error rate by family size. Read family size is indicated on the horizontal axis. The vertical axis gives the mean sequencing error rate for reads with a given family size. Black bars indicate the 95% confidence interval.



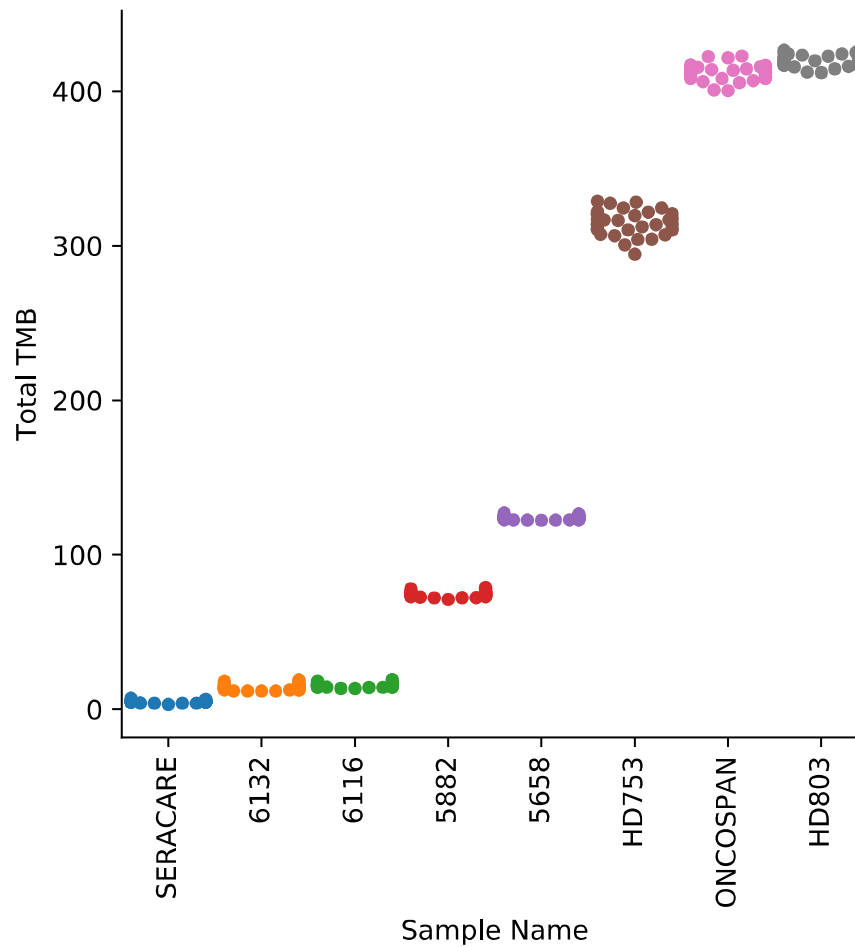
Supplemental Figure 4. MSI scores for normal FFPE samples stratified by ethnicity. The horizontal axis gives the ethnicity of 140 normal FFPE samples assayed using TSO500. The vertical axis gives the MSI score calculated by the TSO500 bioinformatics workflow. The dashed line indicates the MSI score cutoff for determining MSI-high samples.

TruSight Oncology 500

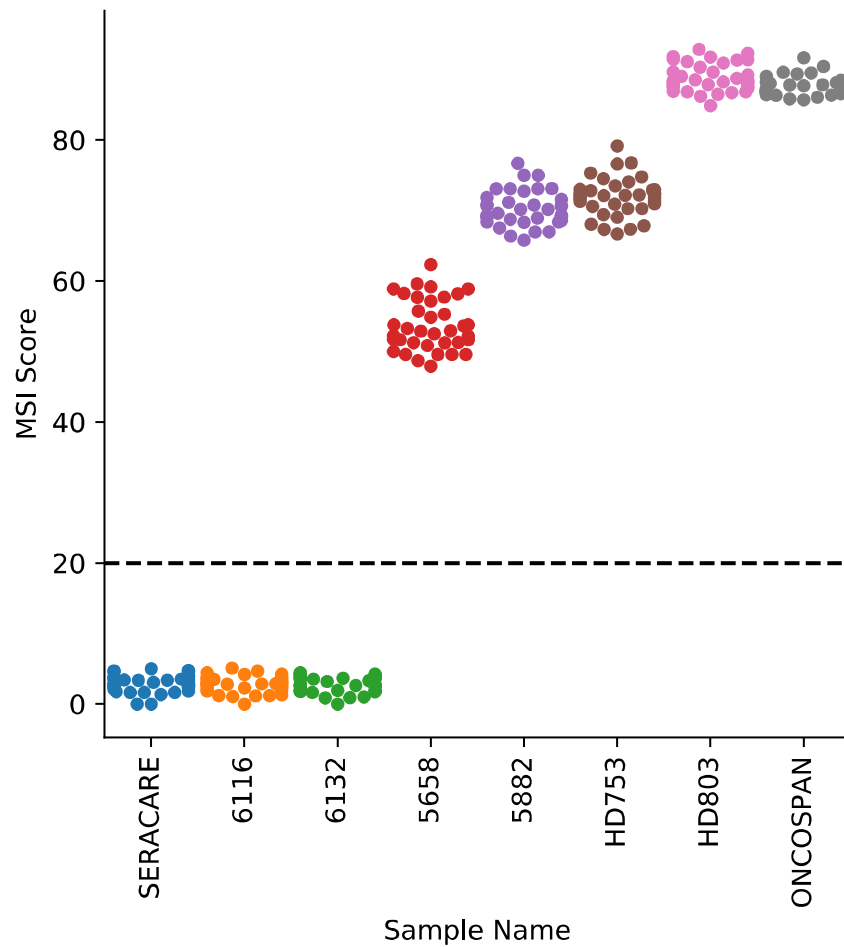


Supplemental Figure 5. Comparison of TSO500 tumor-only TMB values to tumor-normal TMB values (zoomed in to TMB < 50).

TruSight Oncology 500

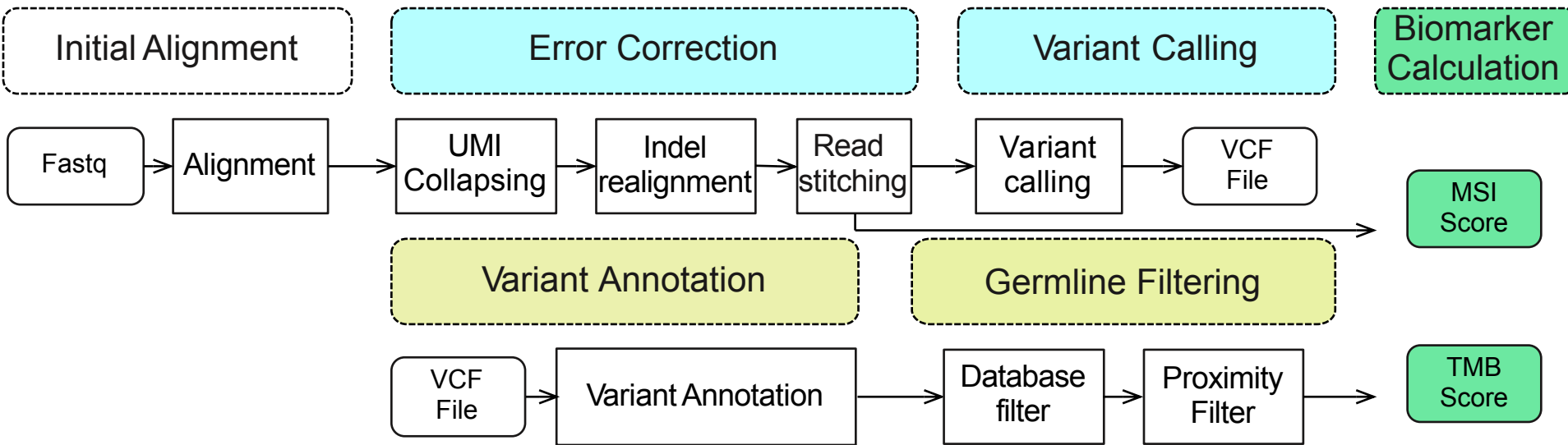


Supplemental Figure 6. Reproducibility of TSO500 tumor-only TMB scores. Tumor-only TMB scores for 36 technical replicates of FFPE lung (6116, 6132), FFPE colon (5882, 5658), and commercial control (HD803, Oncospan, HD753, Seracare) samples are shown.



Supplemental Figure 7. Reproducibility of TSO500 tumor-only MSI scores. Tumor-only MSI scores for 36 technical replicates of FFPE lung (6116, 6132), FFPE colon (5882, 5658), and commercial control (HD803, Oncospan, HD753, Seracare) samples are shown. The dashed line indicates the MSI score cutoff for determining MSI-high samples.

Technical Noise Suppression



Germline Noise Suppression

