



Development of a Novel Reference Material for Tumor Mutational Burden Measurement Based on CRISPR/Cas9 Technology

OPEN ACCESS

Rongxue Peng, Guigao Lin, Lin Li and Jinming Li*

National Center for Clinical Laboratories, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital/National Center of Gerontology, Beijing, China

Edited by:

Alessio Biagioni,
University of Florence, Italy

Reviewed by:

Fengbiao Mao,
Peking University Third Hospital, China
Manuel Pires Bicho,
University of Lisbon, Portugal

*Correspondence:

Jinming Li
jml@nccl.org.cn

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 30 December 2021

Accepted: 08 April 2022

Published: 28 April 2022

Citation:

Peng R, Lin G, Li L and Li J (2022)
Development of a Novel Reference
Material for Tumor Mutational
Burden Measurement Based on
CRISPR/Cas9 Technology.
Front. Oncol. 12:845636.
doi: 10.3389/fonc.2022.845636

As a biomarker that affects treatment decisions of immune checkpoint inhibitors, the accuracy, reliability, and comparability of tumor mutational burden (TMB) estimation is of paramount importance. To improve the consistency and reliability of these tests, qualified reference materials providing ground-truth data are crucial. In this study, we developed a set of formalin-fixed and paraffin-embedded (FFPE) samples with different TMB values as the novel reference materials for TMB estimation. By introducing several clinically relevant variants in MutS Homolog 2 (*MSH2*) gene and DNA polymerase epsilon (*POLE*) gene into human cell lines using CRISPR/Cas9 technology, we first constructed four typical cell lines which verified with hypermutator or ultramutator phenotype. Followed by cell mixing and paraffin embedding, the novel FFPE samples were prepared. It was confirmed that our novel FFPE samples have sufficient quantity of cells, high reproducibility, and they can provide matched wild type sample as the genetic background. The double-platform whole exome sequencing validation showed that our FFPE samples were also highly flexible as they containing different TMB values spanning a clinically relevant range (2.0–106.1 mut/Mb). Without limitations on production and TMB values, our novel FFPE samples based on CRISPR/Cas9 editing are suitable as candidate reference materials. From a practical point of view, these samples can be used for the validation, verification, internal quality control, and proficiency testing of TMB assessment.

Keywords: CRISPR/Cas9, tumor mutational burden, next-generation sequencing, reference material, standardization

INTRODUCTION

Tumor mutational burden (TMB) is the total number of non-synonymous somatic mutations in the genomic coding area (1–4). In theory, the TMB level can reflect the probability of tumor neoantigen production and therefore, the likelihood of immune recognition and tumor cell killing (5). In recent years, numerous studies have confirmed that TMB, characterized by the number of somatic mutations derived from next-generation sequencing (NGS) approaches, can be used as a promising biomarker to predict the efficacy of immune checkpoint inhibitors (6–10). Hence, based on the abundant research data, guidelines from the European Society for Medical Oncology and the National Comprehensive Cancer Network have incorporated TMB as a biomarker in non-small cell lung cancer and gastric cancer in succession, recommending the combination of ipilimumab plus nivolumab or pembrolizumab as a first-line or second-line treatment for patients with high TMB (11–13).

As a biomarker that affects treatment decisions, the accuracy, reliability, and comparability of TMB estimation is of paramount importance (14–17). Currently, TMB is typically detected by whole exome sequencing (WES) and comprehensive genomic profiling. A series of U.S. Food and Drug Administration approved kits and laboratory-developed tests have been applied in clinical practice (18–20). However, as the testing panels, sequencing platforms, and bioinformatic algorithms differ widely across assays, and the mutation types considered for TMB estimation can vary from one laboratory to another, significant differences in TMB levels were always noticed between different assays and laboratories, especially when TMB values were around levels that may be clinical decision points (14, 15, 17, 21). Although many attempts have been made in recent years to improve the measurement of TMB, the inconsistency of testing is still an important problem that is yet to be resolved (14, 15, 17, 21). For clinical laboratories to improve the consistency and reliability of these tests, qualified reference materials that provide ground-truth data in harmonizing these measurements are of prime importance (21).

To date, the materials typically used are formalin-fixed and paraffin-embedded (FFPE) surgical specimens or paraffin-embedded human-derived tumor cell lines (with or without matched normal cell lines) (21, 22). However, owing to the difficulty of obtaining and mass-producing these materials, intratumor and intertumor heterogeneity (23, 24), and variations in the TMB levels in different patients (6, 25), FFPE samples from cancer patients display poor reproducibility. Meanwhile, since the validated TMB level in most human-derived tumor-normal matched cell lines that are publicly available and well-characterized is always lower than 10 mutations/megabase (mut/Mb) (22, 26), the applicability of tumor-normal cell line-derived samples is hampered by the

limited number of cell lines with high TMB values and the cell types. For FFPE within tumor-only cell line samples, in spite of the diverse TMB values and cell types, they are not suitable for all NGS assays because they cannot provide a consistent genomic background. As such, it is obvious that these materials are not perfect as a standard for stringently assessing the performance of TMB estimation. To address these problems, a novel reference material should be developed. Since there is a high degree of variation in TMB levels across different cancer types (25), the ideal reference materials should be widely available, highly reproducible, and flexible enough to generate a series of standards with different TMB levels in ways that are applicable to different situations.

As many studies have demonstrated that mutations in the exonuclease domain of mismatch repair (MMR) genes and DNA polymerase epsilon (*POLE*) gene, which take charge of DNA repair in high-fidelity DNA replication, can cause a hypermutator or ultramutator phenotype (27–29). Hence, in this study, we developed a CRISPR/Cas9 knock-in system to introduce several clinically relevant variants in the MutS Homolog 2 (*MSH2*) and *POLE* gene into human cell lines, and constructed a series of cell lines with hypermutator or ultramutator phenotype. Followed by mixing with wild type cells at a range of precise ratios and paraffin embedding, thus, we developed a series of novel FFPE reference materials for TMB analysis. Without limitations on TMB level and production, these samples have a sufficient quantity of cells with high reproducibility, and they can be applied to any existing TMB estimation assay, since they can provide matched wild type sample as a genetic background. Moreover, apart from the FFPE sample, our reference material can be further prepared into any sample type, including circulating tumor DNA.

MATERIALS AND METHOD

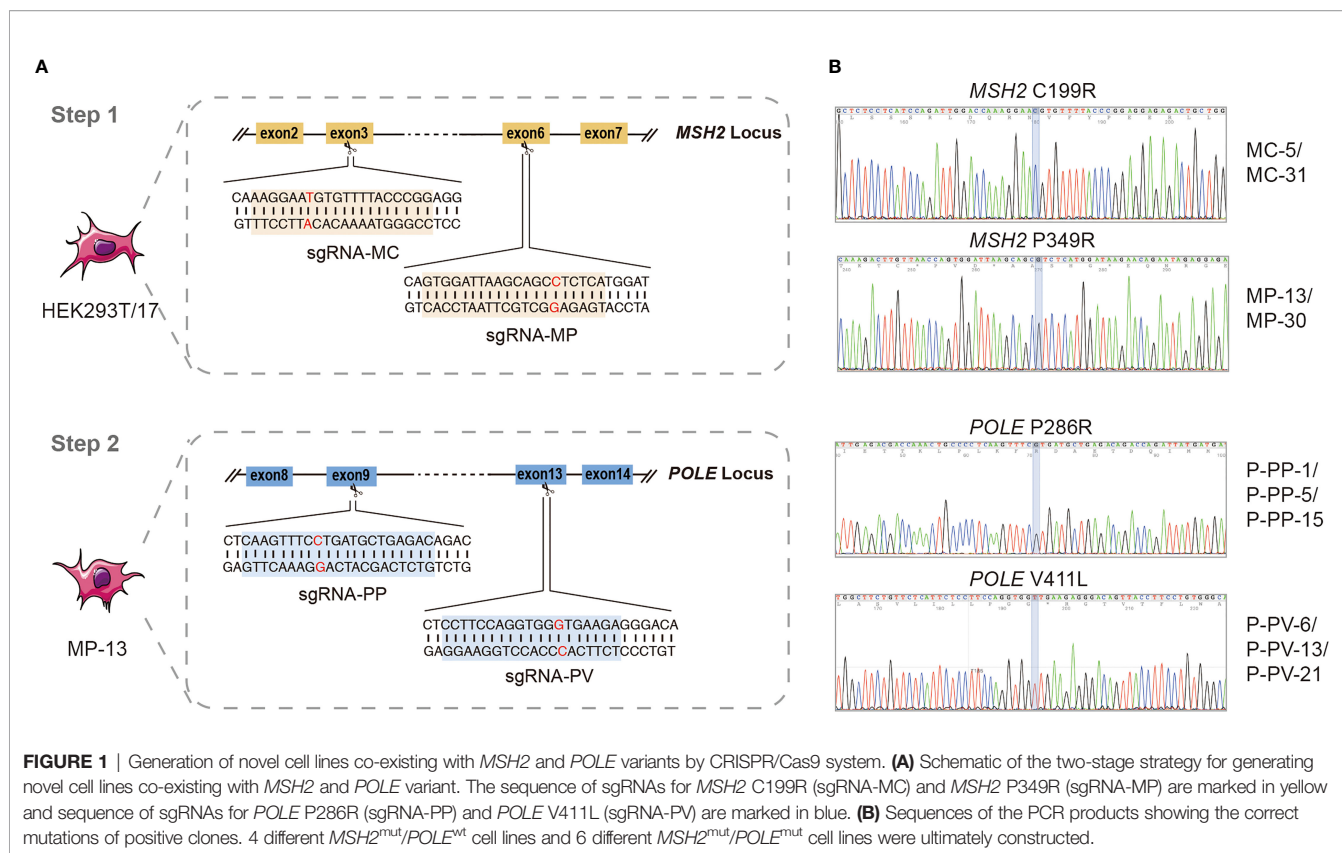
Cell Line and Culture Conditions

HEK293T/17 cell line was chosen in our experiments as it is easy to culture and transfect. The cell line was purchased from American Type Culture Collection (ATCC, Manassas, USA), and cultured in Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific, Waltham, USA) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific, Waltham, USA) at 37°C and 5% CO₂.

Generation of Novel Cell Lines Co-Existing With *MSH2* and *POLE* Variants

CRISPR/Cas9 gene editing technology was used to generate novel cell lines co-existing with *MSH2* and *POLE* variants in our study. The process of gene editing was fundamentally divided into two stages: cell lines containing *MSH2* variants was first generated by editing HEK293T/17 cells and then the cell lines containing the *MSH2* variant were further edited in order to generate cell lines co-existing with *MSH2* and *POLE* variants (Figure 1).

Abbreviations: TMB, tumor mutational burden; FFPE, formalin-fixed and paraffin-embedded; WES, whole exome sequencing; NGS, next-generation sequencing; MMR, mismatch repair; sgRNA, single guide RNA; SpCas9, Streptococcus pyogenes Cas9; ssODN, single strand deoxyoligonucleotide; VAF, variant allele frequency; HDR, homology directed repair; LUAD, Lung adenocarcinoma.



For each *MSH2* and *POLE* variant, single guide RNA (sgRNA) in close proximity to the site of mutation (less than 10 nt) was firstly designed *in silico* (<http://crispor.tefor.net/>; last access date: December 5, 2021). Subsequently, a suitable sgRNA was manually identified using the following criteria: a NGG motif on the 3' end of sgRNA, a G (guanine) appended on the 5' end of sgRNA, and a high on-target specificity and low off-target probability as possible (Table 1) (30, 31). DNA oligos containing the sgRNA sequences were synthesized and constructed into the Px458 vector (purchased from ATCC, Manassas, USA) expressing Cas9 endonuclease and green fluorescent protein.

For each transfection, well-dissociated cells at a density of 0.5×10^6 /mL were transfected using an electroporation kit (Celextrix LLC, Manassas, USA) on Celextrix LE+ machine according to the manufacturer's instructions. 6 μ g of sequence-verified plasmid (Px458 backbone) expressing specific sgRNA and *Streptococcus pyogenes* Cas9 (SpCas9) were transfected along with 3 μ L of 100 μ M single-stranded deoxyoligonucleotide (ssODN) (90 nt with 5' PO₄ and 3' OH modification) (Sangon, Shanghai, China) carrying the mutation to be incorporated. After electroporation, the cells were plated onto 12-well plates with pre-warmed medium for further culturing. Simultaneously, SCR-7 (Selleck, Shanghai, China) was added at a final concentration of 1 μ M for 48 h, with the aim of increasing the rate of homology directed repair (HDR). After 48 h, the transfection efficiency was assessed by observing the fluorescence intensity of green fluorescent protein, and then, the mixed clonal cells were digested for further investigation. A fraction

of the cells was used to extract genomic DNA for evaluating the cleavage efficiencies of different sgRNAs by PCR and Sanger sequencing, whereas the remaining cells were flow cytometrically sorted with a SONY SH800 cell sorter (SONY Biotechnology, California, USA) into 96-well plates (Corning, Corning, NY) for monoclonal cell isolation and culture. Once the single-cell derived clones reached approximately 1×10^6 , the cells with CRISPR/Cas9-edited homozygous variants were rapidly screened by PCR and Sanger sequencing. The primers used for PCR are listed in Table 1.

WES Validation for CRISPR/Cas9 Edited Cell Lines

Positive clones from the Sanger sequencing screen were subsequently detected by WES to ascertain the mutation spectra and TMB levels.

Briefly, genomic DNA was extracted using the QIAmp DNA Mini Kit (QIAGEN, Hilden, Germany) from the cell lines grown for 5–10 passages after editing. 400 ng extracted DNA was used for sequencing library preparation. By using a combination of HiSeq NGS Ultima DNA Library Prep Kit for MGI (YEASEN, Shanghai, China) and 96rxn xGEN-lockdown-reagents (IDT, Coralville, USA), the sequencing libraries were prepared according to the manufacturer's instructions. Sequencing was performed using 2×150 bp paired-end reads on DNBSQ-T7RS (MGI, Shenzhen, China) with an average depth of $500 \times$ according to standard sequencing protocols. Data analysis was performed using a custom-processed pipeline. First, the raw

TABLE 1 | Sequences of sgRNAs and primers for each MSH2 and POLE variant.

Target Site	sgRNA	ssODN	Forward primer to amplify target sites	Reverse primer to amplify target sites
MSH2 C199R	5'-GCCGGGTAAACACATTCCCTT (AGG)-3'	5'-ATCTTGAGGCTCCTCATTCCAGATTGGACCAAGGAACGTTGTTTA CCCGAGGAGAGACTGCTGGAGACATGGGAAACTGAGACAGA -3'	5'-TCAGTTTGAAGACATTTCTTTGG -3'	5'-TCACTAGACTCAATTTGGTTACCTG -3'
MSH2 P349R	5'-GTGGATTAAGCAGCCTCTCA (TGG)-3'	5'-AAAACCCCTCAAGGACAAAGACTTTAACACAGTGGATTAAAGCAGCGT CTCATGGATAAGAACAGAAATAGAGGAGAGATTGAAATTTAGTG -3'	5'-TTTGGCTGGGGGAGAAATGT -3'	5'-GGTATAATCATGTGGGTAACCTGC -3'
POLE P286R	5'-GTCTCAGCATCAGGAAACTTG (AGG)-3'	5'-ATCGATCATGTAGGAAATCATCAATCTGGTCTGCTCAGCATCAC GAAACTTGAGGGGAGTTTGGTCGCTCAATGTCAAAATGCCAA -3'	5'-GGGAGTTTAGAGCTTGGCTT -3'	5'-TCCGTTCTTCCCACAAATACCCG -3'
POLE V411L	5'-GCCTTCCAGGTGGGTGAAGA (GGG)-3'	5'-AGAATCATCTGGTCTGTCTCTCATTCTCCTCCAGGTGGTTGAAG AGGCAGATTACCTTCCCTGTGGCAGTCATATCTCAAGGGGG -3'	5'-GGTGCCTTCTTGTGTGGTGG -3'	5'-GTCTGAGGAGAGAACGCCAG -3'

Red "C": for single guide RNA (sgRNA) not beginning with Guanine, an extra G was added to facilitate efficient transcription from U6 promoter; Green Base: altering variant.

reads were filtered to remove adapters and low-quality reads using fastp (version 0.20.1). Subsequently, alignment sorting, local realignment, duplicate removal, and base quality recalibration were performed using Sentieon (version 202010). Mutect2 (version 4.1.9) and SomVAS (an in-house software) were used to call and annotate the SNVs and indels.

The variant-calling criteria for TMB calculation were set as allelic depth ≥ 3 , variant allele frequency (VAF) $\geq 5\%$, nonsynonymous mutations, and located in the coding regions. For the TMB calculation, the number of genomic alterations after the filtering process was divided by 32 Mb. R packages tidyverse, maftools, and deconstructSigs were used to analyze the mutation data.

Preparation of FFPE Reference Materials

Confluent cell cultures were digested, and counted by Invitrogen Countess 3 FL (Thermo Fisher Scientific, Waltham, USA) prior to mixing and processing. Subsequently, according to the expected TMB level by calculating, cells with ultramutator phenotype were serially blended with the HEK293T/17 cells at precise ratios of 1:2, 1:3, 1:4, 1:5, 1:6, and 1:7, with total cells of 1×10^8 . The mixed cells were pelleted by centrifugation, and the medium was carefully aspirated. To fix cell pellets and construct a spatial network structure to agglomerate cells without affecting the cell shape, a cell block preparation kit (LBP, Guangzhou, China) was used. After 2 h, the cell pellet was carefully dislodged, wrapped in paper, and placed in processing cassettes. Dehydrating and paraffin embedding were subsequently automated performed using a Renaissance Tissue Processor (Ventana Medical Systems, Tucson, USA). Each paraffin-embedded cell block was cut into 10 μm -thick sections and used as the FFPE reference material for TMB analysis.

WES Validation for FFPE Reference Materials

In order to estimate the TMB level of our novel FFPE reference materials, the genomic DNA extracted from 10 μm -thick FFPE slides using the GeneRead DNA FFPE Kit (QIAGEN, Hilden, Germany) was subjected to two reference laboratories for WES analysis (performed on DNBSEQ-T7RS [MGI, Shenzhen, China] and Novaseq 6000 [Illumina, Santiago, USA], respectively). Both assays have been strictly validated and have passed our multicenter assessment (unpublished data). The WES assay performed on DNBSEQ-T7RS was performed as described above, while the WES assay performed on Novaseq 6000 was processed as follows.

200ng DNA was sent for WES performed on Illumina Novaseq 6000 (2x150 bp) with libraries prepared using the Human Exome Library Preparation kit (Genetronhealth, Beijing, China) and sequenced to an average depth of 500x according to the manufacturer's instructions. After removing adapters and low-quality reads by using FastQC (version 0.11.5) and Trimmoatic (version 0.36), the raw reads were aligned to the human genome (hg38) using BWA (version 0.7.17-r1194). Samtools (version 1.3), Picard (version 2.2.1), and GATK (version 3.5) were sequentially used for alignment sorting, duplicate removal, and base quality

recalibration. Variant calling was performed using Mutect (version 3.1.0-g72492bb) and Strelka (version 2.9.2). An in-house software, Result_gather, was used to retain only high-confidence mutations for subsequent TMB estimation: VAF < 5%, allelic depth < 7, and dbSNPs were all removed. For further TMB calculations, tmbTissue.py (in-house software) was used.

Data and Statistical Analysis

The mutation data of lung adenocarcinoma (LUAD) were retrieved from The Cancer Genome Atlas database (<https://portal.gdc.cancer.gov/legacy-archive>) for mutational signatures analysis and comparison. WES data used for commutability assessment derived from our unpublished research and the Sequence Read Archive under accessions PRJNA307199 (32) and PRJNA293912 (33). Statistical analysis and data visualization were processed by R packages tidyverse, maftools, deconstructSigs, and GraphPad Prism 8.

RESULTS

Generation of Novel Cell Lines Co-Existing With *MSH2* and *POLE* Variants Using the CRISPR/Cas9 System

In order to affect the cellular DNA repair function as well as generate cell line with hypermutator or ultramutator phenotype, two previously deemed Class 5 pathogenic *MSH2* variants (*MSH2* C199R and *MSH2* P349R) and two previously deemed Class 5 pathogenic *POLE* variants (*POLE* P286R and *POLE* V411L) were selected. A two-step CRISPR/Cas9 gene editing strategy, as outlined in **Figure 1**, was used to sequentially introduce each variant directly into the endogenous *MSH2* or *POLE* loci.

The first step of the genome editing process was to introduce variants directly into the endogenous *MSH2* locus of HEK293T/17 cells (**Figure 1A**). To do this, a plasmid vector-based expression of both sgRNA and SpCas9 was used to introduce a genomic DNA double-strand break at the appropriate location in *MSH2*. We then selected clones which successfully utilized homology directed repair (HDR) to site-specifically introduce the mutations in a homozygous fashion. To enhance the generation efficiency of single-cell derived clones carrying the desired mutation, SCR-7 (DNA ligase IV inhibitor) was used during the targeting process to reduce non-homologous end joining repair of DNA double-strand break and increase the HDR efficiency. As determined by Sanger sequencing screening, we identified homozygous clones with efficiencies ranging from 3.7–5.1%. Ultimately, four different *MSH2* homozygous variant-expressing cell lines were successfully generated in the first-round gene editing, including MC-5 and MC-31 for *MSH2*^{C199R/C199R} cell lines, and MP-13 and MP-30 for *MSH2*^{P349R/P349R} cell lines (**Figure 1B**).

The second step of the editing process began with the creation of the *MSH2* homozygous variant-expressing cell line (MP-13) described above (**Figure 1A**). A similar gene editing protocol was used to introduce variants directly into the endogenous *POLE*

locus of MP-13. To select individual positive clones, we performed fluorescence-activated cell sorting, and eventually identified six novel cell lines co-existing with *MSH2* and *POLE* variants (P-PP-1, P-PP-5, and P-PP-15 for *MSH2*^{P349R/P349R}/*POLE*^{P286R/P286R} cell lines; P-PV-6, P-PV-13, and P-PV-21 for *MSH2*^{P349R/P349R}/*POLE*^{V411L/V411L}) by sequencing the specific *MSH2* and *POLE* loci (**Figure 1B**). Higher editing efficiencies ranging from 14.3–15% were observed in secondary editing.

Mutation Landscape and TMB Level of the Edited Cells

To ascertain the mutation landscape and TMB levels of each kind of edited cell, validation was performed for MC-5, MP-13, P-PP-1, and P-PV-6 as representatives. Using WES, which is the gold standard for TMB estimation, all samples were sequenced with an average coverage depth of more than 500×, and all target mutations had sufficient coverage to accurately detect variants with VAF as low as 5%. The wild-type HEK293T/17 cell line was also detected in parallel as a normal control to filter out irrelevant mutations and measure the TMB level of edited cell lines.

Results obtained by WES revealed that, compared to the HEK293T/17 cell line, there was a clear increase of mutations in all these edited cell lines (**Figure 2A**). Of which, the number of detected somatic mutations in P-PP-1, and P-PV-6 (*MSH2*^{mut}/*POLE*^{mut}) was significantly higher than that in MC-5, MP-13 (*MSH2*^{mut}/*POLE*^{wt}). According to statistics, mutagenesis has respectively increased 1.22–1.25-fold and 8.31–16.8-fold in *MSH2*^{mut}/*POLE*^{wt} cells and *MSH2*^{mut}/*POLE*^{mut} cells over *MSH2*^{wt}/*POLE*^{wt} control cells after editing. In all edited cells, the newly generated mutations randomly distributed in all chromosomes, both introns and exons, with VAF ranging from 2–99% (**Figure 2B**).

For further exploring the differences between these engineered cell lines and authentic tumors, we then analyzed the mutational pattern and signature of nascent mutations in these four edited cell lines. It was found that, just as the mutational pattern of most tumors, the bulk of nascent mutations in our edited cells were SNVs (72.8–96.6%), with the remainder consisting of deletions and insertions (Indels, 2.5–20.4%) and Del-Ins (**Figure 3**). In MC-5 and MP-13 (*MSH2*^{mut}/*POLE*^{wt}), missense mutations were the most frequent alterations (65.99–71.54%), followed by frameshift Indels (23.58–24.49%), nonsense mutations (2.44–3.40%), and splice-site mutations (2.44–2.72%); whereas for P-PP-1 and P-PV-6 (*MSH2*^{mut}/*POLE*^{mut}), missense mutations mainly posed 87.84–90.58% of nascent somatic mutations (**Figure 3**). Almost half of this observed increase in P-PP-1 and P-PV-6 was due to mutations bearing strong similarity to those identified in tumors with concurrent DNA replication infidelity and MMR deficiency (34–36). These include C>A transversions and C>T transitions with a strong preference for NCC motifs. The signature of the remaining increased nascent mutations was also observed in MC-5 and MP-13, with a clear increase in C>T transversions in the CCG trinucleotide context, which is consistent with DNA damage arising from culturing cells *in vitro* (**Figure 4A**) (37, 38). Moreover, by analyzing the similarities of the mutational

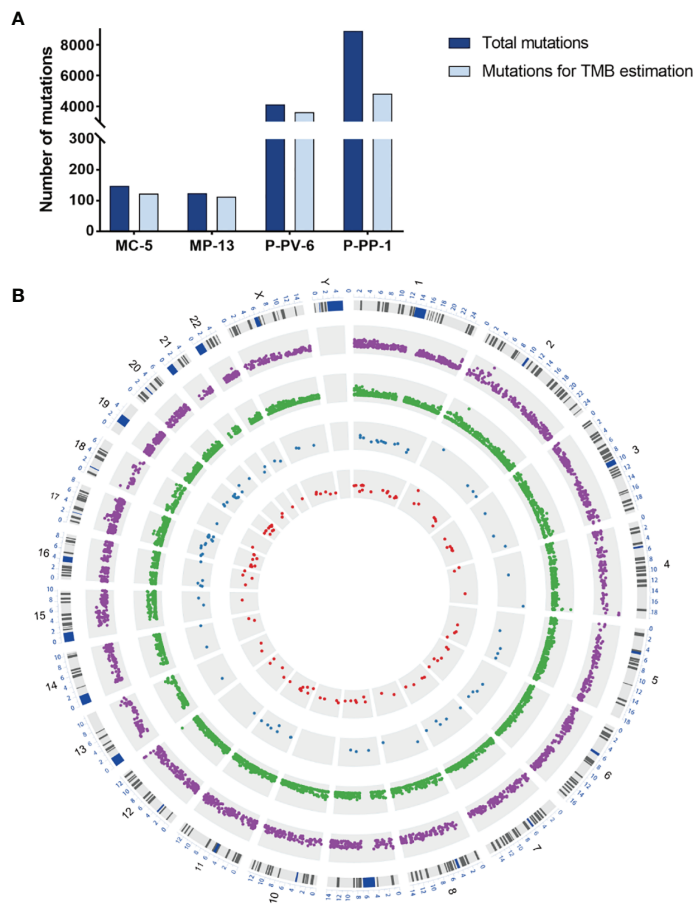


FIGURE 2 | The mutation landscape and TMB levels of MC-5, MP-13, P-PP-1, and P-PV-6. **(A)** Distribution of total nascent mutations and nonsynonymous mutations in four engineered cell lines. **(B)** Circos Plot of detected nascent mutations in four engineered cell lines. Outer purple Track: P-PV-6. Middle green track: P-PP-1. Middle blue track: MP-13. Inner red track: MC-5.

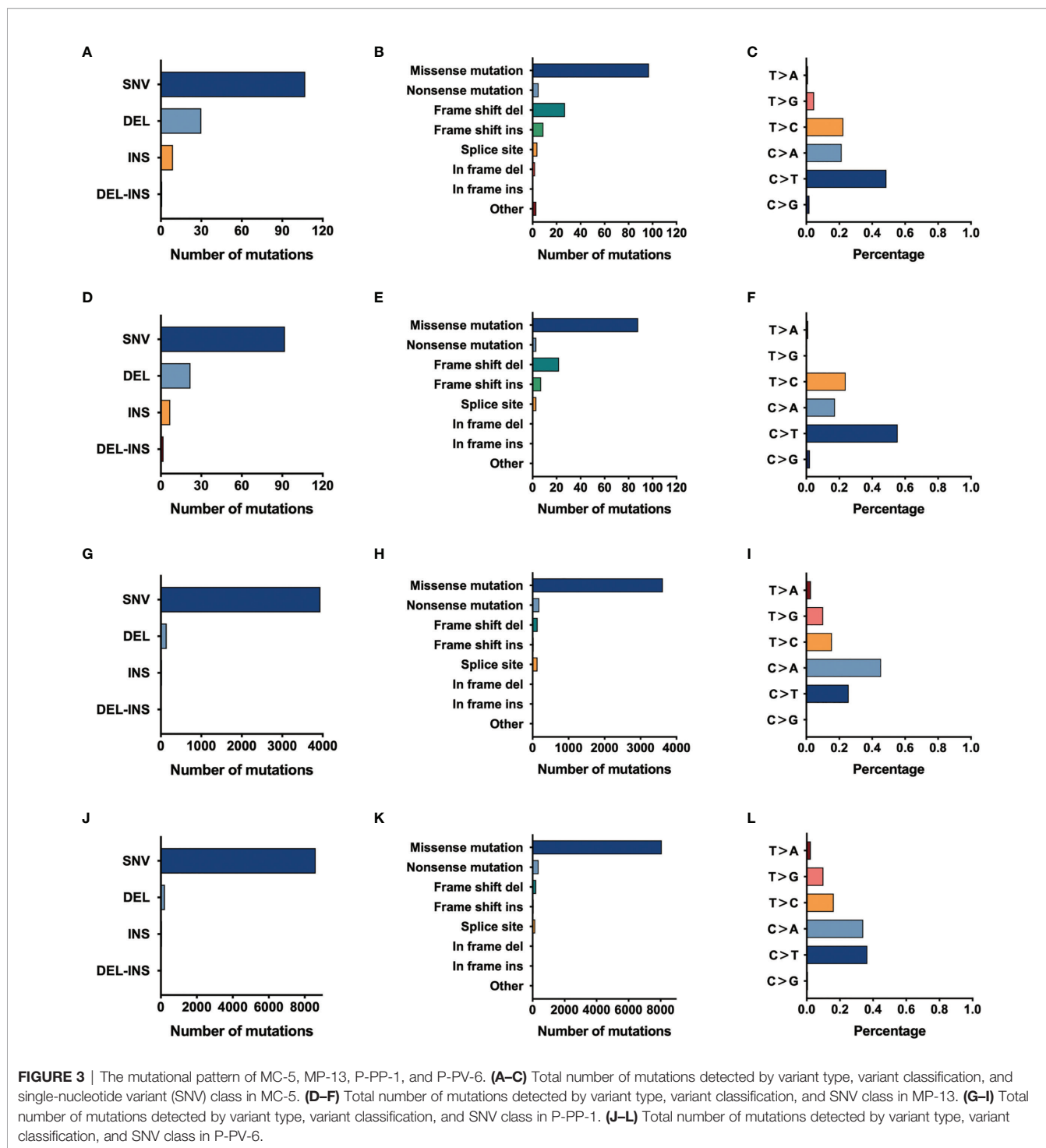
signatures, our edited cells also displayed features similar to those observed in LUAD with high TMB (**Figure 4B**).

By further filtering nonsynonymous mutations located in the coding regions, the calculated TMB levels for MC-5, MP-13, P-PP-1, and P-PV-6 were 3.8 (123 muts), 3.5 (113 muts), 113.3 (3626 muts), and 151.0 mut/Mb (4832 muts), respectively. It was found that although our edited cell lines will accumulate more mutations after several of passages, these nascent mutations during passages have little impact on TMB levels as most of the nascent mutations are located in non-coding regions. Take P-PP-1 as an example, although a total of 790 mutations have raised after 5 passages (from P5 to P10, approximately 15 generations), however, the calculated TMB levels climbed only 1.5 mut/Mb, as 113.3 mut/Mb (3626 muts) for P-PP-1 (P5) and 114.8 mut/Mb (3672 muts) for P-PP-1 (P10). Similarly, the calculated TMB levels of MP-13 raised from 2.6 mut/Mb (83 muts) to 3.5 mut/Mb (113 muts) after 8 passages (from P2 to P10, approximately 24 generations). Hence, it was demonstrated that these edited cells are feasible enough to prepare reference materials with specific TMB level.

Characteristics and Validation Results of the FFPE Reference Materials

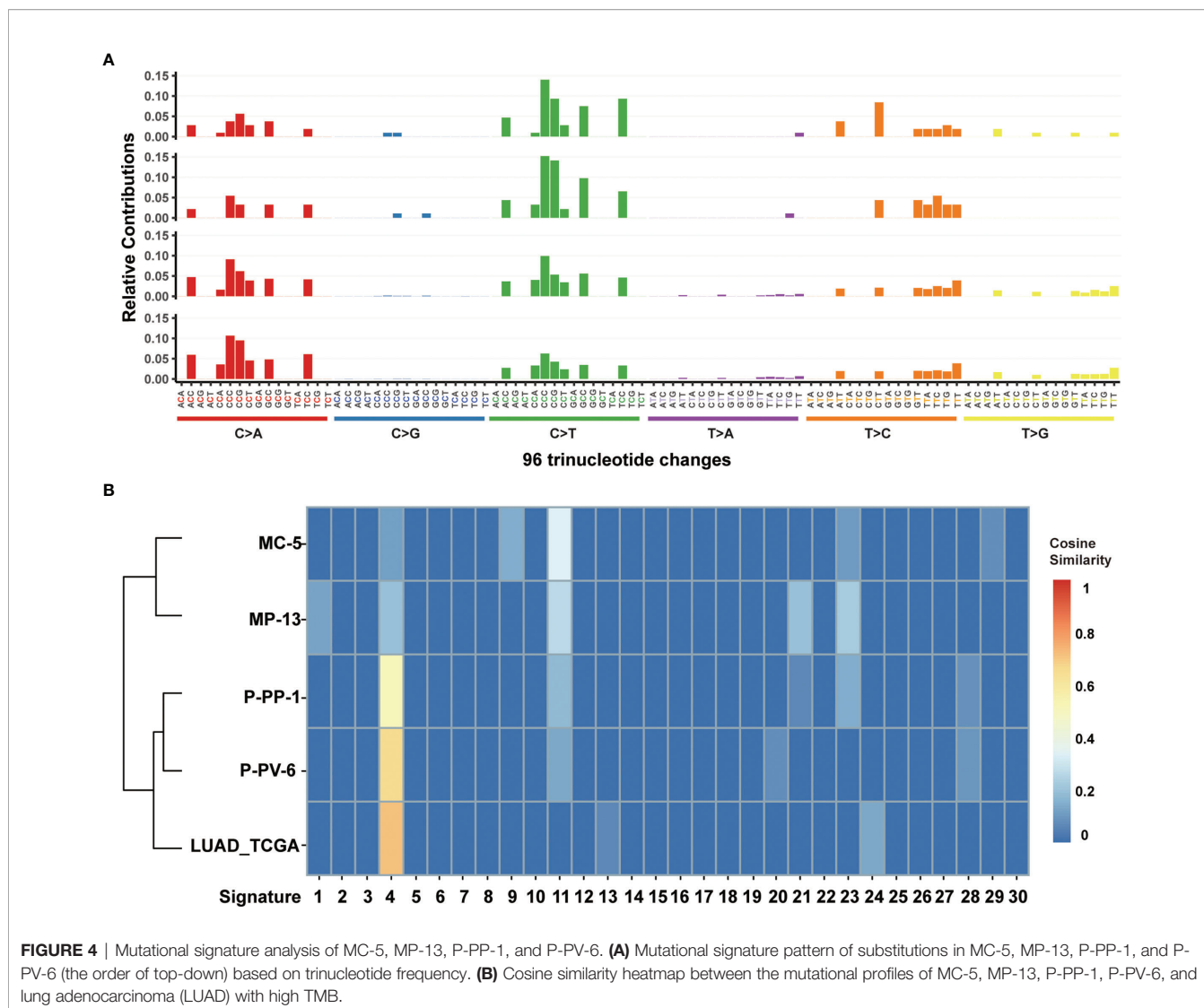
Based on genomic alterations characterized by WES, P-PP-1 was selected to further prepare FFPE reference materials with a gradient of TMB levels as it has an ultramutator phenotype. With a total of 1×10^8 cells, P-PP-1 were serially blended with the HEK293T/17 cells at a range of precise ratios of 1:2, 1:3, 1:4, 1:5, 1:6, and 1:7 (named as P:H 1:2, P:H 1:3, P:H 1:4, P:H 1:5, P:H 1:6, P:H 1:7, respectively). After fixing, each cell pellet was embedded into small blocks as about 5–10 mm long, 5–10 mm wide, and 3–4 mm high. Besides, HEK293T/17 cells were also paraffin-embedded and applied as paired normal samples. To determine whether the amount of total DNA was sufficient to meet the requirements of subsequent detection, each FFPE cell block was sliced and subsequently extracted DNA. The total amount of genomic DNA in each section was higher than 1 μ g, which was adequate for the detection of WES and comprehensive genomic profiling.

In order to further identify the feasibility, homogeneity, and stability of our novel FFPE reference materials as well as



determine the TMB level, a sample validation was carried out. Sections were randomly sampled from each type of FFPE block, and detected by two different WES assays performed on DNBSEQ-T7RS and Novaseq 6000 as previously described in the Methods section. With an average coverage depth of more than 500×, as well as sufficient coverage to accurately detect variants with VAF as low as 5%, the TMB values were obtained from each test, and the average of the two assays was used as the

definite TMB level for the prepared reference materials. Depending on statistics, although the calculated TMB values differed slightly between different assays, the average results of P-PP-1 derived FFPE samples were basically in accordance with the TMB value detected by cells, as 106.1 ± 8.6 mut/Mb. The results of FFPE reference materials derived from gradient diluted P-PP-1 were also consistent with the theoretical results, spanning a clinically relevant range (2.0–46.1 mut/Mb), which partly reflects



the homogeneity of our novel FFPE reference materials. The detailed results are shown in **Figure 5**. Besides, the stability assessment showed that no significant changes were observed in the amount of extracted DNA of our novel FFPE samples after storage at 4°C for 2 months. These results confirmed the feasibility, homogeneity, and stability of our novel FFPE reference materials.

Furthermore, in order to compare the commutability of our samples with the real clinical samples, we further compared the WES data of clinical samples which derived from previous published studies with those of our reference materials. The quality control statistics of the raw sequencing data showed that our FFPE samples-derived reads were similar to reads derived from clinical specimens in the sequencing quality pattern. The overall alignment rate of reads and average target capture efficiency of our novel FFPE samples were also approximating to those of the clinical specimens (overall alignment rate: 99.94% vs 99.96%; average target capture efficiency: 53.34% vs 60.64%; **Table 2**). Besides, we also investigated the mutational pattern of

variants included in TMB calculation in our samples. The results showed that our samples contained different types of clinically relevant variations, ranging from SNVs to indels, just as the mutational pattern of clinical specimens. Hence, our novel FFPE samples based on CRISPR/Cas9 editing were confirmed with good commutability and availability as candidate reference materials for TMB estimation.

DISCUSSION

In the emerging clinical paradigm of ICI therapy, accurate detection is important to make the right decision for the treatment of cancer patients, for instance, in the case of TMB assessment for selecting the subgroup of patients receiving ipilimumab or pembrolizumab therapy (1). Among the numerous factors that affect the accuracy of TMB measurements, the lack of proper reference materials is particularly prominent (15, 17, 21). Therefore, to ensure the

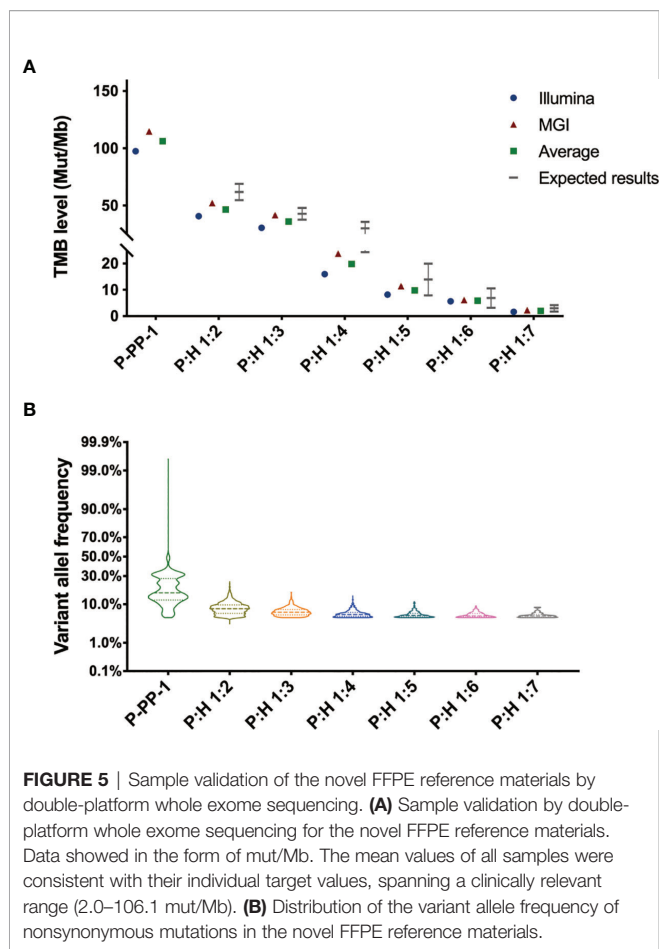


FIGURE 5 | Sample validation of the novel FFPE reference materials by double-platform whole exome sequencing. **(A)** Sample validation by double-platform whole exome sequencing for the novel FFPE reference materials. Data showed in the form of mut/Mb. The mean values of all samples were consistent with their individual target values, spanning a clinically relevant range (2.0–106.1 mut/Mb). **(B)** Distribution of the variant allele frequency of nonsynonymous mutations in the novel FFPE reference materials.

accuracy and reproducibility of TMB measurements, the present study sought to develop a kind of well-characterized candidate reference material.

As many studies have demonstrated that mutations affecting MMR and replicative DNA polymerases can cause rapid mutagenesis and a significant increase in TMB level (28, 34, 38), therefore in our study, we generated several cell lines with different TMB levels by introducing homozygous cancer-associated *MSH2* and *POLE* variants through the use of

CRISPR/Cas9 system, and hereby developed a series of novel FFPE reference materials for TMB analysis by paraffin embedding these constructed cell lines. By using a two-stage gene editing process, four different *MSH2^{mut}/POLE^{wt}* cell lines and six different *MSH2^{mut}/POLE^{mut}* cell lines were finally constructed and named as MC-5, MC-31, MP-13, MP-30, P-PP-1, P-PP-5, P-PP-15, P-PV-6, P-PV-13, and P-PV-21, respectively (**Figure 1**). Of these, MC-5, MP-13, P-PP-1, and P-PV-6 were randomly selected as representative samples, and verified by WES to ascertain the mutation landscape and TMB levels of each kind of edited cell. Compared to the wild-type HEK293T/17 cells, a significant increase of mutations and TMB values was observed in these engineered cells, especially in *MSH2^{mut}/POLE^{mut}* cells (P-PP-1 and P-PV-6, **Figure 2**). The mutational signature of these cells bore strong similarity to those identified in LUAD with high TMB, which confirms the availability of our edited cell lines (**Figures 3, 4**). Besides, the relatively stable TMB level of these cells during passages also demonstrate the feasibility of the edited cell lines. After subsequent cell mixing, a series of cell pellets with different TMB estimates were obtained and embedded as FFPE blocks. With a sufficient amount of genomic DNA, our FFPE samples were detected by two different WES assays. The results indicated that our FFPE samples were homogenous and highly flexible as they contained different TMB values spanning a clinically relevant range (2.0–106.1 mut/Mb, **Figure 5**). Hence, we verified the suitability of our novel FFPE samples as reference materials for TMB assessment in clinical practice.

In contrast to the paraffin blocks of tumor tissues, our novel FFPE reference materials have proven to be sustainable and homogeneous. Since the cell lines can be grown *in vitro*, these edited cells and wild-type cells are potentially available in large quantities and can be mixed thoroughly, which makes this material easy to obtain and homogenize. In addition, unlike the FFPE within tumor-only cell line samples, our novel FFPE reference materials can generate paired variational and normal samples with the identical genetic background of HEK293T/17 cells. By sequence alignment with the genetic background, single nucleotide polymorphisms and germline mutations can be sufficiently filtered, and somatic mutations can be accurately identified by any existing NGS assays for TMB assessment. Not

TABLE 2 | Quality control statistics and mutational pattern comparison between the novel reference materials and real clinical samples.

Sample	Quality Parameters						Variant Type				TMB Value	
	Total reads	Reads mapped	Percent of reads on target	Mean targeted coverage(De-dup)	Target efficiency	Q30	SNV	del	ins	del-ins	Muts	Muts/Mb
P-PP-1	639790136	99.94%	57.50%	492.31	54.20%	96.14%	2866	84	15	0	2965	97.53
P:H 1:2	665295500	99.94%	55.90%	517.50	52.95%	96.24%	1226	10	2	0	1238	40.72
P:H 1:3	836595300	99.94%	58.01%	591.87	53.97%	96.34%	926	2	2	0	930	30.59
P:H 1:4	759738783	99.94%	56.05%	537.97	52.23%	96.53%	484	0	1	0	485	15.95
P:H 1:5	770632598	99.92%	54.50%	539.07	50.76%	95.83%	248	0	0	0	248	8.16
P:H 1:6	805576765	99.93%	58.18%	590.83	54.56%	96.08%	170	0	1	0	171	5.63
P:H 1:7	789148550	99.93%	58.49%	569.01	54.71%	96.12%	50	0	0	0	50	1.64
Patient 1	340621864	99.97%	55.93%	377.53	55.86%	97.67%	331	14	0	0	345	11.35
Patient 2	251488840	99.97%	68.65%	307.24	68.11%	97.40%	142	2	0	0	144	4.74
Patient 3	253431462	99.95%	60.26%	218.02	57.95%	97.59%	105	1	1	0	107	3.52

only that, our reference material panel can provide a series of control materials with precise TMB values spanning clinically relevant ranges, especially around levels that may be clinical decision points. Currently, many researches have shown that there is a high degree of variation in TMB levels across different cancer types, and their corresponding clinically applicable TMB threshold for therapy effectiveness of immune checkpoint inhibitors may also different (27, 39). Hence, the critical range controls (with TMB values ranging from 5 mut/Mb to 20 mut/Mb) are greatly needed for laboratories to validate and harmonize TMB-based assays. The edited cells bearing a hypermutation phenotype can be serially blended with the wild-type HEK293T/17 cells at a range of precise ratios to prepare panels in ways that are applicable for different cancers and situations. Furthermore, expect to provide reference materials for TMB values evaluation, our reference materials can also be used to evaluate the consistency and repeatability of somatic mutation detection since our serial FFPE samples were gradient diluted from the same cells. In other words, the laboratory can evaluate the detection consistency by comparing the detected mutations and their VAF in samples with different dilutions. For instance, mutations detected in P:H 1:7 samples should also be detected in P:H 1:6–P:H 1:2 samples, and with higher VAF values. Given all that, our reference materials can validate laboratory-developed tests and verify commercial detection kits. Meanwhile, since our standards are sustainable, homogeneous, and suitable for all existing TMB measurement, it is possible to perform internal quality control and proficiency testing of clinical laboratories that are using various assays by simply using our novel reference materials.

From a practical point of view, our edited cells can further be blended with other edited cells in the future, such as our previously edited cells containing the epidermal growth factor receptor gene (*EGFR*), *KRAS* proto-oncogene, GTPase gene (*KRAS*), and/or the echinoderm microtubule associated protein like 4 and ALK receptor tyrosine kinase (*EML4-ALK*) fusions (40–42). By this kind of mixture, we can simulate the authentic heterogeneous tumors. And as the mixed materials containing both numerous passenger mutations and driver mutations, they are applicable to validate the comprehensive genomic profiling tests as one-time used, thus avoiding repetitive and comprehensive validations. Apart from these, with the new cell lines generated by CRISPR/Cas9 system, circulating tumor DNA

reference materials can also be generated by enzymatic digestion (43).

In this study, we have evaluated a type of novel FFPE sample as a candidate for reference materials for TMB assessment. The FFPE reference materials have been found to be widely available, highly reproducible, and flexible enough to generate a series of standards with different TMB levels in ways that are applicable for different situations.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

RP and JL contributed to conception and design of the study. RP, GL, and LL participated in the acquisition of data and implemented the analysis. RP and JL contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by National Natural Science Foundation of China (No. 81902145; RP), and Beijing Municipal Natural Science Foundation (No. 7204299; RP). The funding sources have no role in the design and conduct of the study, collection, management, analysis, and interpretation of the data, preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

The authors would like to thank Geneplus Technology Inc. and Genetron Technology Inc. for their technical support.

REFERENCES

- Fumet JD, Truntzer C, Yarchoan M, Ghiringhelli F. Tumour Mutational Burden as a Biomarker for Immunotherapy: Current Data and Emerging Concepts. *Eur J Cancer* (2020) 131:40–50. doi: 10.1016/j.ejca.2020.02.038
- Galuppini F, Dal Pozzo CA, Deckert J, Loupakis F, Fassan M, Baffa R. Tumor Mutation Burden: From Comprehensive Mutational Screening to the Clinic. *Cancer Cell Int* (2019) 19:209. doi: 10.1186/s12935-019-0929-4
- Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of Tumor Mutation Burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495
- Li R, Han D, Shi J, Han Y, Tan P, Zhang R, et al. Choosing Tumor Mutational Burden Wisely for Immunotherapy: A Hard Road to Explore. *Biochim Biophys Acta Rev Cancer* (2020) 1874(2):188420. doi: 10.1016/j.bbcan.2020.188420
- Havel JJ, Chowell D, Chan TA. The Evolving Landscape of Biomarkers for Checkpoint Inhibitor Immunotherapy. *Nat Rev Cancer* (2019) 19(3):133–50. doi: 10.1038/s41568-019-0116-x
- Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to Pd-1 Inhibition. *N Engl J Med* (2017) 377(25):2500–1. doi: 10.1056/NEJMc1713444
- Klemptner SJ, Fabrizio D, Bane S, Reinhart M, Peoples T, Ali SM, et al. Tumor Mutational Burden as a Predictive Biomarker for Response to Immune

- Checkpoint Inhibitors: A Review of Current Evidence. *Oncologist* (2020) 25 (1):e147–59. doi: 10.1634/theoncologist.2019-0244
8. Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, et al. Safety, Efficacy and Tumor Mutational Burden as a Biomarker of Overall Survival Benefit in Chemo-Refractory Gastric Cancer Treated With Toripalimab, a Pd-1 Antibody in Phase Ib/Ii Clinical Trial Nct02915432. *Ann Oncol* (2019) 30 (9):1479–86. doi: 10.1093/annonc/mdz197
 9. Ott PA, Bang YJ, Piha-Paul SA, Razak ARA, Bennouna J, Soria JC, et al. T-Cell-Inflamed Gene-Expression Profile, Programmed Death Ligand 1 Expression, and Tumor Mutational Burden Predict Efficacy in Patients Treated With Pembrolizumab Across 20 Cancers: Keynote-028. *J Clin Oncol* (2019) 37(4):318–27. doi: 10.1200/JCO.2018.78.2276
 10. 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. *N Engl J Med* (2018) 378(22):2093–104. doi: 10.1056/NEJMoal1801946
 11. NCCN. *The Nccn Clinical Practice Guidelines in Oncology - Non Small Cell Lung Cancer (Version 1.2019)* (2019). Available at: <http://www.nccn.org>.
 12. Committee EG. *Metastatic Non-Small-Cell Lung Cancer: Esmo Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up* (2020). Available at: <https://www.esmo.org/guidelines>.
 13. Committee EG. *Esmo Clinical Practice Guidelines: Gastrointestinal Cancers* (2019). Available at: <https://www.esmo.org/guidelines>.
 14. Merino DM, McShane LM, Fabrizio D, Funari V, Chen SJ, White JR, et al. Establishing Guidelines to Harmonize Tumor Mutational Burden (Tmb): *In Silico* Assessment of Variation in Tmb Quantification Across Diagnostic Platforms: Phase I of the Friends of Cancer Research Tmb Harmonization Project. *J Immunother Cancer* (2020) 8(1):e000147. doi: 10.1136/jitc-2019-000147
 15. Stenzinger A, Endris V, Budczies J, Merkelbach-Bruse S, Kazdal D, Dietmaier W, et al. Harmonization and Standardization of Panel-Based Tumor Mutational Burden Measurement: Real-World Results and Recommendations of the Quality in Pathology Study. *J Thorac Oncol* (2020) 15(7):1177–89. doi: 10.1016/j.jtho.2020.01.023
 16. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. Esmo Recommendations on Microsatellite Instability Testing for Immunotherapy in Cancer, and Its Relationship With Pd-1/Pd-L1 Expression and Tumour Mutational Burden: A Systematic Review-Based Approach. *Ann Oncol* (2019) 30(8):1232–43. doi: 10.1093/annonc/mdz116
 17. Stenzinger A, Allen JD, Maas J, Stewart MD, Merino DM, Wempe MM, et al. Tumor Mutational Burden Standardization Initiatives: Recommendations for Consistent Tumor Mutational Burden Assessment in Clinical Samples to Guide Immunotherapy Treatment Decisions. *Genes Chromosomes Cancer* (2019) 58(8):578–88. doi: 10.1002/gcc.22733
 18. U.S. Food and Drug Administration. *Foundation One Cdx-P170019/S013: Summary of Safety and Effectiveness Data (Ssed)* (2020). Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019B.pdf.
 19. U.S. Food and Drug Administration. *Evaluation of Automatic Class Iii Designation for Msk-Impact (Integrated Mutation Profiling Of actionable cancer targets)* (2020). Available at: https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170058.pdf.
 20. U.S. Food and Drug Administration. *Pgdx Elio Tissue Complete: 510(K) Substantial Equivalence Determination Decision Summary* (2020). Available at: https://www.accessdata.fda.gov/cdrh_docs/reviews/K192063.pdf.
 21. Vega DM, Yee LM, McShane LM, Williams PM, Chen L, Vilimas T, et al. Aligning Tumor Mutational Burden (Tmb) Quantification Across Diagnostic Platforms: Phase Ii of the Friends of Cancer Research Tmb Harmonization Project. *Ann Oncol* (2021) 32(12):1626–36. doi: 10.1016/j.annonc.2021.09.016
 22. Park S, Lee C, Ku BM, Kim M, Park WY, Kim NKD, et al. Paired Analysis of Tumor Mutation Burden Calculated by Targeted Deep Sequencing Panel and Whole Exome Sequencing in Non-Small Cell Lung Cancer. *BMB Rep* (2021) 54(7):386–91. doi: 10.5483/BMBRep.2021.54.7.045
 23. Fridland S, Choi J, Nam M, Schellenberg SJ, Kim E, Lee G, et al. Assessing Tumor Heterogeneity: Integrating Tissue and Circulating Tumor DNA (Ctdna) Analysis in the Era of Immuno-Oncology - Blood Tmb Is Not the Same as Tissue Tmb. *J Immunother Cancer* (2021) 9(8):e002551. doi: 10.1136/jitc-2021-002551
 24. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, et al. Local Mutational Diversity Drives Intratumoral Immune Heterogeneity in Non-Small Cell Lung Cancer. *Nat Commun* (2018) 9(1):5361. doi: 10.1038/s41467-018-07767-w
 25. Sholl LM, Hirsch FR, Hwang D, Botling J, Lopez-Rios F, Bubendorf L, et al. The Promises and Challenges of Tumor Mutation Burden as an Immunotherapy Biomarker: A Perspective From the International Association for the Study of Lung Cancer Pathology Committee. *J Thorac Oncol* (2020) 15(9):1409–24. doi: 10.1016/j.jtho.2020.05.019
 26. Fang LT, Zhu B, Zhao Y, Chen W, Yang Z, Kerrigan L, et al. Establishing Community Reference Samples, Data and Call Sets for Benchmarking Cancer Mutation Detection Using Whole-Genome Sequencing. *Nat Biotechnol* (2021) 39(9):1151–60. doi: 10.1038/s41587-021-00993-6
 27. Buttner R, Longshore JW, Lopez-Rios F, Merkelbach-Bruse S, Normanno N, Rouleau E, et al. Implementing Tmb Measurement in Clinical Practice: Considerations on Assay Requirements. *Esmo Open* (2019) 4(1):e000442. doi: 10.1136/esmoopen-2018-000442
 28. Hodel KP, Sun MJS, Ungerleider N, Park VS, Williams LG, Bauer DL, et al. Pole Mutation Spectra Are Shaped by the Mutant Allele Identity, Its Abundance, and Mismatch Repair Status. *Mol Cell* (2020) 78(6):1166–77e6. doi: 10.1016/j.molcel.2020.05.012
 29. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* (2017) 171(5):1042–56 e10. doi: 10.1016/j.cell.2017.09.048
 30. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the Crispr-Cas9 System. *Science* (2014) 343(6166):80–4. doi: 10.1126/science.1246981
 31. Peng R, Lin G, Li J. Potential Pitfalls of Crispr/Cas9-Mediated Genome Editing. *FEBS J* (2016) 283(7):1218–31. doi: 10.1111/febs.13586
 32. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and Transcriptomic Features of Response to Anti-Pd-1 Therapy in Metastatic Melanoma. *Cell* (2016) 165(1):35–44. doi: 10.1016/j.cell.2016.02.065
 33. Rizvi NA, Hellmann MD, Snyder A, Kvisiborg P, Makarov V, Havel JJ, et al. Cancer Immunology. Mutational Landscape Determines Sensitivity to Pd-1 Blockade in Non-Small Cell Lung Cancer. *Science* (2015) 348(6230):124–8. doi: 10.1126/science.aaa1348
 34. Li HD, Lu C, Zhang H, Hu Q, Zhang J, Cuevas IC, et al. A Pole p286r Mouse Model of Endometrial Cancer Recapitulates High Mutational Burden and Immunotherapy Response. *JCI Insight* (2020) 5(14):e138829. doi: 10.1172/jci.insight.138829
 35. Haradhvala NJ, Kim J, Maruvka YE, Polak P, Rosebrock D, Livitz D, et al. Distinct Mutational Signatures Characterize Concurrent Loss of Polymerase Proofreading and Mismatch Repair. *Nat Commun* (2018) 9(1):1746. doi: 10.1038/s41467-018-04002-4
 36. Alexandrov LB, Stratton MR. Mutational Signatures: The Patterns of Somatic Mutations Hidden in Cancer Genomes. *Curr Opin Genet Dev* (2014) 24:52–60. doi: 10.1016/j.gde.2013.11.014
 37. Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational Signatures: Emerging Concepts, Caveats and Clinical Applications. *Nat Rev Cancer* (2021) 21(10):619–37. doi: 10.1038/s41568-021-00377-7
 38. Nemeth E, Lovrics A, Gervai JZ, Seki M, Rospo G, Bardelli A, et al. Two Main Mutational Processes Operate in the Absence of DNA Mismatch Repair. *DNA Repair (Amst)* (2020) 89:102827. doi: 10.1016/j.dnarep.2020.102827
 39. Hsiehchen D, Espinoza M, Valero C, Ahn C, Morris LGT. Impact of Tumor Mutational Burden on Checkpoint Inhibitor Drug Eligibility and Outcomes Across Racial Groups. *J Immunother Cancer* (2021) 9(11):e003683. doi: 10.1136/jitc-2021-003683
 40. Lin G, Zhang K, Peng R, Han Y, Xie J, Li J. Clustered Regularly Interspaced Short Palindromic Repeat (Crispr)/Crispr-Associated Endonuclease Cas9-Mediated Homology-Independent Integration for Generating Quality Control Materials for Clinical Molecular Genetic Testing. *J Mol Diagn* (2018) 20(3):373–80. doi: 10.1016/j.jmoldx.2018.02.002
 41. Jia S, Zhang R, Lin G, Peng R, Gao P, Han Y, et al. A Novel Cell Line Generated Using the Crispr/Cas9 Technology as Universal Quality Control Material for Kras G12v Mutation Testing. *J Clin Lab Anal* (2018) 32(5):e22391. doi: 10.1002/jcla.22391

42. Peng R, Zhang R, Lin G, Yang X, Li Z, Zhang K, et al. Crispr/Cas9 Technology-Based Xenograft Tumors as Candidate Reference Materials for Multiple Eml4-Alk Rearrangements Testing. *J Mol Diagn* (2017) 19(5):766–75. doi: 10.1016/j.jmoldx.2017.06.003
43. Zhang R, Peng R, Li Z, Gao P, Jia S, Yang X, et al. Synthetic Circulating Cell-Free DNA as Quality Control Materials for Somatic Mutation Detection in Liquid Biopsy for Cancer. *Clin Chem* (2017) 63(9):1465–75. doi: 10.1373/clinchem.2017.272559

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Peng, Lin, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.