High-Throughput Detection of Clinically Relevant Mutations in Archived Tumor Samples by Multiplexed PCR and Next-Generation Sequencing

Richard Bourgon², Shan Lu¹, Yibing Yan¹, Mark R. Lackner¹, Weiru Wang³, Victor Weigman⁶, David Wang⁵, Yinghui Guan¹, Lisa Ryner¹, Hartmut Koeppen², Rajesh Patel¹, Garret M. Hampton¹, Lukas C. Amler¹, and Yulei Wang¹

Abstract

Purpose: Tailoring cancer treatment to tumor molecular characteristics promises to make personalized medicine a reality. However, reliable genetic profiling of archived clinical specimens has been hindered by limited sensitivity and high false-positive rates. Here, we describe a novel methodology, MMP-seq, which enables sensitive and specific high-throughput, high-content genetic profiling in archived clinical samples.

Experimental Design: We first validated the technical performance of MMP-seq in 66 cancer cell lines and a Latin square cross-dilution of known somatic mutations. We next characterized the performance of MMP-seq in 17 formalin-fixed paraffin-embedded (FFPE) clinical samples using matched fresh-frozen tissue from the same tumors as benchmarks. To demonstrate the potential clinical utility of our methodology, we profiled FFPE tumor samples from 73 patients with endometrial cancer.

Results: We demonstrated that MMP-seq enabled rapid and simultaneous profiling of a panel of 88 cancer genes in 48 samples, and detected variants at frequencies as low as 0.4%. We identified DNA degradation and deamination as the main error sources and developed practical and robust strategies for mitigating these issues, and dramatically reduced the false-positive rate. Applying MMP-seq to a cohort of endometrial tumor samples identified extensive, potentially actionable alterations in the PI3K (phosphoinositide 3-kinase) and RAS pathways, including novel PIK3R1 hotspot mutations that may disrupt negative regulation of PIK3CA.

Conclusions: MMP-seq provides a robust solution for comprehensive, reliable, and high-throughput genetic profiling of clinical tumor samples, paving the way for the incorporation of genomic-based testing into clinical investigation and practice. Clin Cancer Res; 20(8); 2080–91. ©2014 AACR.

Introduction

Comprehensive characterization of mutations in clinically actionable genes and key cancer pathways is necessary to enable personalized approaches to cancer treatment. Although several multiplexed genotyping technologies, for example, primer extension assays, mass spectrometry (MS), and allele-specific PCR (asPCR), have been successfully used in archived clinical samples (1–3), these technologies are limited to assaying well-characterized loci. Next-generation sequencing (NGS) has the potential to overcome these limitations and provide a more comprehensive genetic portrait of individual tumors by permitting de novo variant detection. For example, NGS enables profiling of tumor-suppressor genes, in which loss of function mutations can occur anywhere. Comprehensive profiling is likely to be critical in indications such as endometrial cancer, in which up to 90% of cancer have at least one alteration in the phosphoinositide 3-kinase (PI3K) or RAS pathways, many of which occur at non-hotspot loci, or in tumor-suppressor genes such as PTEN, PIK3R1, and PIK3R2 (4, 5).

Systematic genetic profiling of tumors in the clinical setting presents unique technical and practical challenges. Published studies using NGS methodologies have thus far primarily focused on DNA from large, surgically resected frozen tumor samples (4). In reality, the vast majority of tumor samples are formalin-fixed paraffin-embedded (FFPE) processed. FFPE samples yield DNA in limited quantity and often of low quality due to degradation, cross-linking, and chemical modifications caused by the fixation process. These limitations are
Translational Relevance

Molecular characterization of individual tumors promises to make tailored cancer treatment a reality, but significant challenges exist in deriving reliable genetic profiles from archived clinical tumor samples. Here, we describe a novel methodology, MMP-seq, which enables such profiling with broad content, excellent sensitivity and throughput, and low cost. We provide in-depth characterization of the technical performance of MMP-seq and demonstrate its feasibility and utility in clinical applications. We identified DNA degradation and deamination as the predominant error source in formalin-fixed paraffin-embedded sequencing, and developed robust solutions to overcome these issues, yielding dramatically improved sequencing accuracy. MMP-seq can be implemented in large-scale clinical investigations and provides new opportunities for linking genetic profiles and clinical outcome to targeted cancer therapeutics, ultimately accelerating the development of personalized medicine.

High-Throughput Genomic Profiling of Archived Clinical Tumor Samples

Further compounded by tumor heterogeneity and high background signal from surrounding normal tissue or immune infiltrates. Furthermore, although whole-genome and exome sequencing has provided an extensive catalog of common cancer-associated mutations, it is often still too expensive and labor intensive to be practical for clinical investigations with large sample collections. Finally, the limited per-locus sequencing depth provided by genome-wide surveys may not be sufficient for detection of driver or low-frequency drug-resistance variants in technically challenging clinical FFPE tumor tissues.

To address these challenges, considerable effort has been made to develop and validate PCR-based (6, 7), hybridization-based (8), or other (9, 10) targeted sequencing methods that are applicable to FFPE specimens; several recently published reports have expanded on this work (11–15). However, several limitations or gaps still remain. For example, hybridization-based methods typically require large amounts of tissue and have a more labor-intensive workflow that limits sample throughput; and published PCR-based approaches have been mostly limited to well-characterized hotspot mutations. Furthermore, to address the critical issue of sequencing reliability in real-world clinical samples, an in-depth characterization of sequencing accuracy, error sources, and mitigation strategies is still lacking.

Here, we describe a novel methodology, MMP-seq, which integrates microfluidic multiplex PCR (MMP)-based target enrichment with massively parallel sequencing for high-throughput, high-content genetic profiling of clinically relevant cancer genes. Using cancer cell lines and a dilution series of cell lines containing eight known somatic mutations (Supplementary Table S2). All tissue samples were obtained from commercial sources and had appropriate Institutional Review Board approval. All tumor tissues were subjected to review by a pathologist to confirm diagnosis and tumor content. Macrodissection was performed on FFPE tumor tissue to enrich the tumor percentage to greater than 70%. We also analyzed 69 cancer cell lines (Supplementary Data File S1) and a Latin square design with dilution series of cell lines containing eight known somatic mutations (Supplementary Table S2). Genomic DNA was extracted from FFPE tumor tissue using the QiAamp FFPE Kit (Qiagen) after deparaffinization with Envirene; genomic DNA was extracted from cell lines and fresh-frozen tumors using the QiAamp Kit (Qiagen).

FFPE DNA quality assessment by qPCR-based ruler assay

We developed an FFPE DNA quality control ruler assay to quantify the functional DNA copies in FFPE samples that are suitable for target enrichment by PCR. For this, we leveraged a TaqMan copy number assay for the TRAK2 gene, with an amplicon length of 147 bp (Life Technologies; Assay ID: Hs00911853_cn). We used the standard curve method for the TaqMan real-time PCR Assay (Life Technologies; protocol PN 4397425 Rev. C) to do absolute quantification of functional copy number, on the ViiA7 real-time PCR system. Human blood genomic DNA from normal controls (Roche; #1169112001) was used to prepare the standard curve. Average C, from triplicate measurement of each sample was used for calculating functional DNA copy number.

Treatment of FFPE DNA with UDG

For UDG treatment, UDG and UDG buffer (New England BioLabs) were directly added to DNA
suspension buffer (TEKNOVA) in sample 96-well plates. Reaction plates were incubated at 37 °C for 30 minutes for UDG treatment, followed by 95 °C for 5 minutes, for inactivation. The PCR master mix was then added into the same wells and followed by the standard thermal cycling protocol for the Fluidigm 48.48 Access Array Integrated Fluidic Circuit (IFC).

Selection of targeted genes and amplicon library design

Gene selection and primer design and pooling strategies are described in the main text. All target-specific primers were created using Fluidigm custom primer design service. Multiplex primer pools (10–12 primer pairs/pool) were generated on the basis of in silico analysis of primer compatibility; considerations included partitioning overlapping amplicons into separate primer pools, minimizing the probability of primer dimer formation, and ensuring similarity of guanine, cytosine (GC) content with a single pool. All primers were validated by single-plex PCR and assessment of PCR products for expected size, on LabChip GX (Caliper Life Sciences). Of note, 98.4% of primer pairs validated at the single-plex level.

MMP target enrichment and barcoded genomic DNA library construction

The 963 amplicons targeting 88 genes were divided into two panels (480 for panel 1 and 483 amplicons for panel 2) for target enrichment by Access Array (Fluidigm). The experiments were performed according to the Multiplex Amplicon Tagging Protocol from the manufacturer (see details in Supplementary Methods). The resulting sequencing-ready amplicon libraries were sequenced (2 × 108 bp) on an Illumina GAIIx (one library/lane; Expression Analysis) or MiSeq sequencer.

Preamplification and whole-genome amplification before target enrichment

For the 480-plex preamplification protocol, primer pairs from each panel were pooled. DNA (75 ng) from this pool was preamplified for 15 cycles according to the manufacturer’s instruction (Fluidigm; ADP29; PN 100-2988 A1). Products were then digested with ExoSAP-IT to destroy unused primers, diluted 10-fold, and finally submitted for target enrichment PCR on the Access Array IFC. For the whole-genome amplification (WGA) protocol, the GenoPlex Complete Whole Genome Amplification Kit (Sigma-Aldrich; Cat# WGA2) was used to amplify 75-ng genomic DNA. This was followed by QiaQuick column purification. Of note, 150 ng of post-WGA DNA was for target enrichment PCR on the Access Array IFC.

Sequencing data alignment and primary variant calling

FASTQs were paired-end aligned to the human genome (hg19) using Burrows-Wheeler Alignment (BWA) (16). To better handle indels, unassembled reads and their mate pairs were submitted for a second alignment using Burrows-Wheeler Aligner’s Smith-Waterman Alignment (BWA-SW) (17). PCR-based target enrichment strongly favors the reference base within the primer region; to correct for this, we trimmed sequence from aligned reads that intersected primers. Pileups were created using SAMtools (18), and a custom variant-caller, mpileup-variants (ExpressionAnalysis, software available upon request), was used to generate variant calls. For MiSeq data produced for assessment of UDG treatment on FFPE samples, Genomic Short-read Nucleotide Alignment Program (GSNAP) was used for alignment and variant tallying instead (19). Variants were annotated for impact on RefSeq transcript models using Ensembl Variant Effect Predictor (20). All sequence data referenced in this article have been deposited at the European Genome-phenome Archive (EGA; http://www.ebi.ac.uk/ega/), which is hosted by the European Bioinformatics Institute (EBI), under accession number EGAS00001000674.

Variant post-processing and final calls

We applied two additional post-processing filters when making final variant calls for paired fresh-frozen/FFPE or endometrial tumor samples. First, we required at least 400 total reads. Second, we applied a more stringent minimum variant read frequency cutoff: for the fresh-frozen/FFPE–paired sample analysis we used 10%; for the endometrial panel, we used 2% for 340 well-characterized hotspot mutations; and 10% for all other candidates. For the endometrial panel, we further restricted analysis to Catalogue of Somatic Mutations in Cancer Database (COSMIC)–annotated candidates plus, for fully tiled genes, two additional classes: (i) nonsense, frameshift, or insertion/deletion events likely to disrupt the protein; and (ii) missense substitutions predicted to be deleterious by both sorting intolerant from tolerant (SIFT) (21) and PolyPhen2 (22).

Allele-specific qPCR genotyping

The allele-specific qPCR genotyping was performed as previously described (2).

Mass spectrometric genotyping

For assessment of Latin square variants, mass spectrometric genotyping was performed with the OncoCarta Panel v1.0 Kit (sequenom; Catalog #10225) according to the manufacturer’s user guide. To validate alterations detected by MMP-Seq that were not included in the commercially available OncoCarta panels, custom primers and probes were designed using the Sequenom online Assay Design Suite 1.0 (https://www.mysequenom.com/Tools). Experiments were performed with the Complete iPLEX Pro Reagent Set (Catalog# 10217). Data were analyzed by Sequenom TyperViewer 4.0 software.

PTEN immunohistochemistry

PTEN immunohistochemistry (IHC) was conducted on FFPE tumor samples as previously described (23). The PTEN antibody was obtained from Cell Signaling Technology (clone 138G6), and the assay was run on the Discovery platform (Ventana). PTEN staining was evaluated using an H-score method, to account for heterogeneity of expression: Tumor cells were assigned a score of 0 (no staining), 1 (weak), 2 (moderate), or 3 (strong), and these values were
averaged and multiplied by 100. A score of 0 indicates total absence of PTEN in the tumor compartment; a score of 300 indicates PTEN expression in tumor cells equivalent to surrounding normal and stromal cells.

Results

MMP-based target enrichment for actionable and clinically relevant cancer genes

We selected a panel of 88 clinically relevant cancer genes: genes that encode druggable targets, are clinically actionable, have established diagnostic or prognostic value, or have a high reported mutation frequency in the COSMIC (Supplementary Table S1). Two PCR primer design strategies were applied (Supplementary Fig. S2). For the 15 tumor-suppressor genes, all exons and splicing junctions in each transcript variant were tiled with interleaved amplicons (70 bp average overlap). PIK3CA was also tiled, because many somatic mutations in this gene fall outside known hotspot locations in endometrial cancer (5). For 72 other cancer genes with well-characterized somatic alterations, amplicons covered mutation hotspots only. Altogether, 963 amplicons covered regions of interest (ROI) totaling approximately 150 kb and, including 268 exons, 340 mutation hotspots (Supplementary Data File S3), and more than 10,000 COSMIC-annotated mutations. To ensure optimal target enrichment in degraded FFPE samples, 89% of amplicons were under 150 bp and all amplicons were under 200 bp.

For high-throughput target enrichment, we used the Fluidigm Access Array System (Fluidigm Inc.), which uses a microfluidic technology to simultaneously apply 48 distinct PCR reactions to 48 samples. Amplicons were divided into two panels with 10 to 12 PCR primer pairs per reaction to simultaneously amplify 480 (panel 1) or 483 (panel 2) distinct regions on a single Access Array. For each sample, PCR products were pooled, and a sample-specific 10-base DNA barcode plus sequencer-specific adapter was introduced in a second PCR reaction (Supplementary Fig. S1). This workflow generated 48 barcoded, sequencer-ready libraries per run, suitable for pooling and sequencing in a single reaction.

Highly specific and reproducible target enrichment enabled highly sensitive variant detection

We first assessed the target enrichment and variant detection accuracy of MMP-seq in 66 cancer cell lines (Supplementary File S1) and a Latin square cross-dilution of seven lines with mutually exclusive somatic mutations (Supplementary Table S2).

Barcoded libraries were prepared using 50-ng DNA per panel and sequenced on an Illumina GAIIx (Materials and Methods). On average, samples generated 1.4 million reads, with 98% mapping to the targeted ROIs. In contrast with standard hybridization-based exome-seq, only approximately 50% of reads map to targeted ROIs (24). Such highly specific and efficient target enrichment coupled with the focused nature of our ROIs produced an average of 2,600 reads per amplicon; furthermore, 93% of amplicons achieved coverage within 5-fold of this average. To evaluate reproducibility, replicate data were generated from a single cell line in independent target enrichment and sequencing reactions. The number of aligned reads per amplicon and detected variant read frequency were highly reproducible ($R^2 = 0.95$ and 0.99, respectively; Fig. 1 and Supplementary Fig. S3).

A Latin square cross-dilution of seven cancer cell lines with mutually exclusive mutations (Supplementary Table S2) was designed to assess sensitivity to low-frequency variants of MMP-seq—from 40% down to 0.2%. For seven single nucleotide variants (SNV) and one 15-bp deletion in EGFR, we observed a highly linear relationship between observed and expected variant read frequency (Fig. 1C). We also assessed our Latin square samples with two other mutation profiling platforms: the Sequenom MassARRAY OncoCarta Panel v1.0, a mass spectrometric technology that interrogates 238 common mutations in 19 cancer-associated genes; and a multiplexed asPCR technology that interrogates 116 mutations in 11 genes (2). The average detection limit for MMP-seq was 1.8%, compared with 6.6% for MassARRAY and 3.2% for asPCR (Table 1).

To further characterize performance for a broader class of mutations, we performed targeted sequencing in 66 fresh cancer cell lines and after post-processing, detected a total of 629 protein altering variants—SNVs and small insertions and deletions (indels). For the 47 cell lines also represented in COSMIC, all 97 previously reported SNVs and 22 of 24 previously reported small indels (<20 bp) were detected in primary variant calls from our study (Supplementary File S6). Using custom-designed mass spectrometric genotyping assays, we validated 25 (93%) of 27 novel and potentially deleterious SNVs found in our data but not annotated in COSMIC (Supplementary File S4). Several of these may potentially affect pathway activation and drug sensitivity.

Characterization and optimization of sequencing performance in FFPE tumor samples

We next characterized the performance of MMP-seq in FFPE clinical samples using paired fresh-frozen/FFPE tissue from 17 patients with cancer (Materials and Methods). Specifically, we used the matched fresh-frozen samples as a benchmark to evaluate sequencing sensitivity and specificity in FFPE samples. This is conservative, because some observed discrepancies may be due to differences in normal contamination fraction (Supplementary File S2) or true subclonal mutations in the tumor (25). Furthermore, to address the impact of variability in FFPE-associated DNA degradation or chemical modifications, we developed a DNA quality ruler assay and characterized its relationship to sequencing accuracy. The ruler assay provides an estimate of the number of functional DNA copies available for target enrichment. The assay is based on qPCR amplification of the TRAK2 locus—chosen because the amplicon length (149 bp) is consistent with our library and copy number variation at this locus is rare (Materials and Methods).

Functional DNA copy number estimated by the ruler assay varied considerably in our 17 FFPE tumor samples—from
200 to 5,800 copies in 150 ng of DNA—and, critically, was highly predictive of concordance between paired FFPE and fresh-frozen samples. Figure 2A illustrates this for two FFPE samples, one with high and one with low estimated functional copy count. Although good concordance was observed for the former, the latter generated a large number of false positives (i.e., called variants not detected in the matched fresh-frozen sample). Importantly, these false negatives resulted in a high number of false positives, as indicated by the scatter plots in Figure 2A.
positivees included both low- (<10%) and high-frequency (>20%) calls. Using all 17 paired fresh-frozen/FFPE samples, we showed that this held true generally: Improved specificity (here computed as positive predictive value, PPV) was strongly associated with increasing functional copies (Fig. 2B and Supplementary Fig. S4). Notably, specificity approached 100% for high-copy FFPE samples when using more stringent variant read frequency cutoffs. On the other hand, the impact of variant read frequency cutoff and the ruler assay was much more limited for variant detection sensitivity. This suggests that low functional copy FFPE samples may still be suitable for genotyping well-characterized hotspot mutations. Indeed, at the 116 hotspots in 11 cancer genes assayed by asPCR (Supplementary File S3), we observed near-perfect concordance between fresh-frozen sequencing, FFPE sequencing, and asPCR across all 17 patients independent of the functional copies of these samples (Table 2).

FFPE fixation deaminates certain bases; most prominently, cytosine deaminates to uracil, which pairs with adenine and mimics a C>T mutation. To assess the impact on our data, we computed a deamination ratio for all putative FFPE variants (Table 2). Figure 2C shows that FFPE-associated deamination was a very significant source of false positives for some samples, and the ruler assay was strongly predictive for this effect (presumably due to reduced primer annealing). Strategies exist for repairing [26] or selecting against [27, 28] deaminated DNA, and such preprocessing steps may yield improved sequencing accuracy for degraded samples.

Indeed, pretreatment of FFPE DNA samples with UDG to remove uracil-containing deaminated DNA molecules resulted in dramatic reduction in false positives compared with those without UDG treatment (Fig. 2E), with an overall reduction of 77%, and 94% of these corresponding to C>T or G>A changes (Supplementary Table S4). On the other hand, UDG treatment showed little impact on sensitivity in all tested samples regardless of their DNA quality (Supplementary Fig. S5).

<table>
<thead>
<tr>
<th>Mutation</th>
<th>MMP-Seq</th>
<th>MassARRAY</th>
<th>asPCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAF V600E</td>
<td>2.5%</td>
<td>12.4%</td>
<td>2.5%</td>
</tr>
<tr>
<td>EGFR T790M</td>
<td>1.1%</td>
<td>1.1%</td>
<td>1.1%</td>
</tr>
<tr>
<td>EGFR L858R</td>
<td>1.6%</td>
<td>3.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>EGFR E746-A750Δ</td>
<td>0.9%</td>
<td>6.9%</td>
<td>0.9%</td>
</tr>
<tr>
<td>KRAS G12D</td>
<td>1.3%</td>
<td>12.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>NRAS Q61L</td>
<td>0.4%</td>
<td>3.3%</td>
<td>3.3%</td>
</tr>
<tr>
<td>PIK3CA E545K</td>
<td>5.8%</td>
<td>5.8%</td>
<td>5.8%</td>
</tr>
<tr>
<td>PIK3CA H1047R</td>
<td>0.5%</td>
<td>7.6%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Average</td>
<td>1.8%</td>
<td>6.6%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

NOTE: Minimum estimated effective variant frequency at which variant was detected, on the basis of cellular mixture proportions and observed variant read frequencies in sequence data from pure source cell lines (Supplementary Table S3). Minimum detectible frequency varied by locus and technology; for all eight mutations, NGS sensitivity matched or exceeded MS (Sequenom OncoCarta v1 MassARRAY) or asPCR (Fluidigm).

Comprehensive genetic profiling in endometrial cancer clinical samples

Activation of the PI3K and RAS pathways is a main driver in endometrial cancer development. This can occur by inactivation of PTEN (29, 30), activating mutations in PIK3CA (31, 32), AKT1 (33), and KRAS (34, 35), and mutations in PIK3R1 and PIK3R2 (36). Several novel therapeutics targeting the PI3K/AKT/mTOR axis are currently in early clinical development for endometrial cancer (36), and mutation profiles for implicated genes provide promising biomarker candidates. However, the frequency and number of distinct alterations puts these beyond the reach of traditional, single-analyte approaches.

To demonstrate the potential clinical utility of our methodology, we evaluated FFPE tumor samples from 80 patients with endometrial cancer (77 with endometrioid histology; Supplementary File S2). To ensure high sensitivity and specificity, we adjusted DNA input (155 to 419 ng/sample) to provide 3,000 to 5,000 functional copies for target enrichment; 73 of the samples (91%) were of sufficient quality to permit this. For these, average total mapped reads per sample (1.2 million), reads per amplicon (2,100), and coverage uniformity (89% within 5-fold of per sample average) were comparable with what was achieved for fresh-frozen samples. Fraction of reads on target (98%) was also comparable with fresh-frozen samples and substantially better than previously reported for hybridization-based targeted sequencing of FFPE material (~40%; ref. 8).

To better understand the relationship between MMP-Seq and lower-throughput approaches, we also genotyped all 73 samples using multiplexed asPCR (Materials and Methods). At the 81 SNVs interrogated by both technologies, 54 of 55 asPCR-detected variants (98%) were also detected by MMP-Seq, and 54 of 82 MMP-Seq variants (66%) were detected by asPCR. Figure 3C shows that most of the variants detected by MMP-Seq but not asPCR were found at low frequency (2%–5%), as expected given MMP-Seq’s higher sensitivity (Fig. 1C; Table 1).

Table 1. Variant detection sensitivity for Latin square cross-dilution samples

<table>
<thead>
<tr>
<th>Mutation</th>
<th>MMP-Seq</th>
<th>MassARRAY</th>
<th>asPCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIK3CA E545K</td>
<td>5.8%</td>
<td>5.8%</td>
<td>5.8%</td>
</tr>
<tr>
<td>PIK3CA H1047R</td>
<td>0.5%</td>
<td>7.6%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Average</td>
<td>1.8%</td>
<td>6.6%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

High-Throughput Genomic Profiling of Archived Clinical Tumor Samples
Among PI3K and RAS pathway genes, PTEN and PIK3CA were the most frequently mutated (63% and 58% of subjects, respectively; Fig. 3A). We observed highly significant overlap between PTEN and PIK3CA mutations ($P = 0.003$) but mutual exclusivity for PTEN and AKT1 ($P = 0.003$). No other pairwise comparisons showed significant cooccurrence or mutual exclusivity after adjusting for multiple testing. We found a high level of agreement between mutation frequencies from our data and the recently published The Cancer Genome Atlas (TCGA) study of 248 fresh-frozen endometrial tumors (Fig. 3B; ref. 4).

Among the detected variants in PTEN, 59% were non-sense mutations, indels, or frameshifts that would likely lead to an incomplete protein and were classified as "detrimental." Notably, a substantial number of these variants are novel (Fig. 4A). To assess their functional relevance, PTEN protein level was assessed by IHC and found to be significantly lower for the detrimental class (Fig. 4B).

---

**Figure 2.** Optimization and characterization of sequencing performance in FFPE tumor tissues. A, correlation of variant read frequency for two FFPE versus fresh-frozen (FF) sample pairs: one with high-quality FFPE DNA (5,763 functional copies) and the other, low (363 functional copies). B, relationship between functional DNA copies and specificity and sensitivity in de novo variant detection. C, relationship between deamination and functional DNA copies for all putative FFPE false positives. D, FFPE-associated deamination is a very significant source of false positives for low-quality FFPE samples (patient 31251 shown). E, false positives in four FFPE DNA samples with varying DNA quality were identified using matched FF samples as a benchmark. False positives associated with deamination (C$\rightarrow$T or G$\rightarrow$A substitutions) were significantly reduced by UDG pretreatment; false positives associated with other substitutions were also slightly reduced for some samples.
Mutations were detected throughout the coding region of PIK3CA, except in the RAS-binding domain (Fig. 4A). The most frequent mutations were R88Q (16%) within the adaptor-binding domain that binds to the regulatory p85 family members encoded by PIK3R1 and PIK3R2 (37); E545K and Q546H or R (18% combined) in the helical domain; and H1047R (7%) in the kinase domain. This distribution is distinct from that observed in other cancer types, in which the vast majority of mutations cluster within the helical and kinase domains (38, 39).

We confirmed frequent mutations in PIK3R1 and PIK3R2 in endometrial cancer, and an enrichment for PIK3R1 mutations in the region between the SH2 domains (iSH2), which binds to the catalytic subunit (C2) of PIK3CA (Fig. 4A and Supplementary Fig. S6; refs. 4, 5, 37). Interestingly, a majority of the mutations found in PIK3R1 (66%) are deleterious, which is consistent with a potential tumor-suppressor role, inhibiting PIK3CA activation. In addition, two amino acids within the PIK3R1 iSH2 domain were mutated in multiple patients: N564D (3 patients) and K567E or Q (4 patients). These residues are at the interface between PIK3R1 and the PIK3CA C2 domain (Fig. 4C). PIK3R1 N564 forms a hydrogen bond with PI3KCA N345 (40), and this interaction is facilitated by PIK3CA V344, which seems to serve as a scaffold to support the orientation of N345 and the associated loop. In our data, mutations at N344 or N345 of PIK3CA do not cooccur with mutations at N564 of PIK3R1, suggesting that alteration of any one of these crucial amino acids is sufficient to disrupt interaction between the proteins and increase PIK3CA kinase activity via reduction of inhibitory contact with PIK3R1.

Discussion

MMP-seq is a novel methodology that integrates MMP-based target enrichment with massively parallel sequencing for high-throughput, high-content targeted sequencing of cancer genes in archived clinical samples. It effectively tackles important technical and practical challenges not addressed by existing methods, and provides a robust solution for comprehensive genetic characterization of FFPE tumor specimens.

The short-amplicon PCR-based target enrichment of MMP-seq achieved excellent specificity, uniformity, and reproducibility for degraded FFPE DNA. It yielded deep coverage of targeted regions, which translated into variant detection sensitivity superior to that of other genotyping technologies. This high sensitivity is especially important for the detection of clinically relevant low-frequency alterations. EGFR T790M, for example, accounts for approximately 50% of acquired resistance to the first generation of EGFR tyrosine kinase inhibitors (e.g., gefitinib or erlotinib) in non–small cell lung cancer (41). Many novel agents (e.g., afatinib or dacomitinib) have been developed to overcome T790M-mediated resistance and have shown encouraging results in preclinical and clinical settings (42). MMP-Seq can detect T790M variant at true frequencies as low as 1.1% (Fig. 1C; Table 1). Such sensitivity could enable

**Table 2. asPCR validation of hotspot mutations detected by MMP-seq in fresh-frozen or FFPE samples**

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Functional copies</th>
<th>asPCR</th>
<th>Fresh-frozen sequencing</th>
<th>Variant frequency</th>
<th>FFPE sequencing</th>
<th>Variant frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>13908</td>
<td>207</td>
<td>PIK3CA H1047R</td>
<td>PIK3CA H1047R</td>
<td>18.8%</td>
<td>PIK3CA H1047R</td>
<td>78.3%</td>
</tr>
<tr>
<td>13948 150 ng</td>
<td>4,155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13948 100 ng</td>
<td>2,770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13957</td>
<td>247</td>
<td>PIK3CA E542K</td>
<td>PIK3CA E542K</td>
<td>41.4%</td>
<td>PIK3CA E542K</td>
<td>19.7%</td>
</tr>
<tr>
<td>14002</td>
<td>359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14044</td>
<td>1,111</td>
<td>PIK3CA H1047R</td>
<td>PIK3CA H1047R</td>
<td>41.8%</td>
<td>PIK3CA H1047R</td>
<td>29.7%</td>
</tr>
<tr>
<td>21642</td>
<td>2,757</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21725</td>
<td>1,419</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21926</td>
<td>1,015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22050</td>
<td>3,733</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31208</td>
<td>2,190</td>
<td>KRAS G12D</td>
<td>KRAS G12D</td>
<td>36.9%</td>
<td>KRAS G12D</td>
<td>38.7%</td>
</tr>
<tr>
<td>31251</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31480</td>
<td>2,011</td>
<td>MET T1010I</td>
<td>MET T1010I</td>
<td>37.4%</td>
<td>MET T1010I</td>
<td>49.3%</td>
</tr>
<tr>
<td>31499</td>
<td>257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31662</td>
<td>1,418</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41256</td>
<td>2,857</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41442</td>
<td>5,006</td>
<td>KRAS G12A</td>
<td>KRAS G12A</td>
<td>33.1%</td>
<td>KRAS G12A</td>
<td>16.1%</td>
</tr>
<tr>
<td>41463</td>
<td>5,763</td>
<td>BRAF V600E</td>
<td>BRAF V600E</td>
<td>19.4%</td>
<td>BRAF V600E</td>
<td>23.1%</td>
</tr>
</tbody>
</table>

NOTE: Empty cells denote no detected mutations at asPCR-interrogated loci. For MMP-seq genotype calls, a 10% variant frequency cutoff was applied. For 10 of the 17 subjects, asPCR was performed on both fresh-frozen and FFPE samples, and in all such cases results were identical for both samples in the pair. All three approaches (asPCR, fresh-frozen sequencing, and FFPE sequencing) make identical calls for all samples—except for patient 31480, which showed no mutations by asPCR but had high-confidence MET T1010I MMP-seq calls in both fresh-frozen and FFPE.
early identification of this emerging resistance allele—in tumor tissue or circulating tumor DNA—and thereby guide treatment.

Multiplexed PCR-based target enrichment coupled with sample barcoding enabled a broader survey of mutations and the analysis of a large number of samples in each sequencing run, permitting efficient and cost-effective genetic profiling in a large clinical trial. Furthermore, the highly automated microfluidic device requires minimal hands-on time, making it ideal for routine use in clinical laboratories.

To guide clinical decisions, it is critical to ensure the quality and reliability of sequencing results. We demonstrated the utility of a qPCR-based functional copy ruler assay for assessing FFPE DNA quality and for predicting sequencing accuracy. More importantly, we showed that errors associated with low functional DNA copy numbers and DNA deamination artifacts account for the vast majority of false positives. This is consistent with a model in which deamination may be common on aggregate, but is a rare random event at a given locus. When starting with a sufficient number of functional copies in the PCR target capture reactions, rare per-locus changes will read out at low frequency in the sequencing results, and these will then be discarded by standard frequency–based filtering strategies. When starting with a very small number of functional DNA copies, on the other hand, stochastic sampling will often cause high-level amplification of rare per-locus changes—producing high variant read frequencies that pass all filters, resulting in a false positive. Use of the ruler assay permits identification of samples with limited numbers of functional DNA copies and for which stochastic sampling effects and frequency misestimation are likely to be a significant concern. Biochemical removal of deaminated DNA with UDG largely eliminated the deamination-associated false positives that this phenomenon creates. It is worth noting, however, that misestimation of variant frequency for real low-frequency changes is likely to still be an issue for poor-quality samples. Caution is, therefore, required when interpreting observed variant read frequencies for samples with a limited number of functional copies.

The observed relationship between FFPE DNA quality and sequencing accuracy suggests a two-tiered strategy for FFPE tumor samples which is gated by DNA quality.
detecting novel mutations—including those predicted to inactivate tumor-suppressor genes—high specificity is critical due to the large number of loci under consideration, and such applications would benefit from higher DNA quality and variant read frequency cutoffs (Fig. 2B). When assessing well-characterized hotspot mutations, on the other hand, false positives are unlikely to occur by chance at the small number loci in question (even if numerous genome wide). In such cases, one might consider evaluating samples with suboptimal DNA quality, and also reducing the variant read frequency cutoff to increase sensitivity.

Preamplification and WGA have previously been applied for mutation detection from minute amounts of clinical sample DNA (6, 43). In our hands, both (Materials and Methods) created significant complications. Preamplification produced poor uniformity of target coverage (not shown) as well as very high rates of discordance with paired fresh-frozen samples: Sensitivity of 28% and PPV of only 26% at the 10% variant frequency threshold used in Fig. 2B. WGA produced uniform of target coverage but a high false-positive rate relative to paired fresh-frozen (PPV 14%), even for high-quality FFPE samples. Taken together, these findings suggest that caution is required when applying these sample preparation techniques.

In terms of clinical applicability, MMP-seq is a significant advance over existing methodologies: It enables comprehensive profiling of FFPE samples, with broader content, better sensitivity, and higher throughput. As illustration, we applied this powerful approach to an endometrial cancer panel and were able to recapitulate and expand upon recently published results obtained from a larger collection of fresh-frozen material, identifying extensive mutations in the PI3K and RAS pathways—both known mutations characteristic to this cancer type as well as novel loss-of-function
alterations. New therapeutics targeting these frequently deregulated pathways are currently under early clinical development. Preclinical studies have shown that biomarkers in these pathways, including PIK3CA and PTEN mutations, correlate with sensitivity to targeted pan-PI3K and mTOR inhibition in endometrial endometrial cell lines (44). Our study suggests an additional mechanism for PI3K pathway activation in endometrial cancer: Release of negative regulation of PIK3CA activity via either point mutations in the interacting PIK3CA (C2) and PIK3R1 (iSH2) domains or the more frequent deleterious mutations. Similar mutations in PIK3R1 have previously been identified in colorectal cancer and shown to abrogate PIK3CA-inhibitory activity, thereby promoting cell survival, Akt activation, and oncogenesis (45). MMP-seq assessment of a potential association between these known and novel genetic alterations and patient response to treatment is currently under investigation in several clinical studies.

In conclusion, this study demonstrates the technical feasibility and clinical applicability of MMP-seq for the comprehensive profiling of actionable genetic alterations in FFPE tumor samples. This approach can be readily implemented in large-scale clinical investigations for retrospective and, potentially, prospective profiling of tumor cohorts; and it provides new opportunities for linking genetic profiles and patient response to existing and novel cancer therapeutics, ultimately accelerating the development of personalized treatment.

References


High-Throughput Genomic Profiling of Archived Clinical Tumor Samples