The Journal of Molecular Diagnostics, Vol. 18, No. 6, November 2016

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags

Astrid Eijkelenboom,* Eveline J. Kamping,† Annemiek W. Kastner-van Raaij,* Sandra J. Hendriks-Cornelissen,* Kornelia Neveling,† Roland P. Kuiper,† Alexander Hoischen,†‡ Marcel R. Nelen,† Marjolijn J.L. Ligtenberg,*† and Bastiaan B.J. Tops*

From the Departments of Pathology and Human Genetics† and the Donders Centre for Neuroscience,‡ Radboud University Medical Center, Nijmegen, the Netherlands*

CME Accreditation Statement: This activity ("JMD 2016 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint providership of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2016 CME Program in Molecular Diagnostics") for a maximum of 36 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication June 8, 2016.

Address correspondence to Bastiaan Tops, Ph.D., Radboud University Medical Center, PO Box 9101, 6500 HB, Nijmegen, the Netherlands. E-mail: bastiaan.tops@radboudumc.nl.

Sequencing of tumor DNA to detect genetic aberrations is becoming increasingly important, not only to refine cancer diagnoses but also to predict response to targeted treatments. Next-generation sequencing is widely adopted in diagnostics for the analyses of DNA extracted from routinely processed formalin-fixed, paraffin-embedded tissue, fine-needle aspirates, or cytologic smears. PCR-based enrichment strategies are usually required to obtain sufficient read depth for reliable detection of genetic aberrations. However, although the read depth relates to sensitivity and specificity, PCR duplicates generated during target enrichment may result in overestimation of library complexity, which may result in false-negative results. Here, we report the validation of a 23-gene panel covering 41 hotspot regions using single-molecule tagging of DNA molecules by single-molecule molecular inversion probes (smMIPs), allowing assessment of library complexity. The smMIP approach outperforms Sanger and Ampliseq-Personal Genome Machine—based sequencing in our clinical diagnostic setting. Furthermore, single-molecule tags allow consensus sequence read formation, allowing detection to 1% allele frequency and reliable exclusion of variants to 3%. The number of false-positive calls is also markedly reduced (>10-fold), and our panel design allows for distinction between true mutations and deamination artifacts. Not only is this technique superior, smMIP-based library preparation is also scalable, easy to automate, and flexible. We have thus implemented this approach for sequence analysis of clinical samples in our routine diagnostic workflow. *(J Mol Diagn 2016, 18: 851—863; http://dx.doi.org/10.1016/j.jmoldx.2016.06.010)*

The availability and requirement of molecular therapeutics in routine cancer treatment has greatly increased over the past decade. Combined with the stratification of patients amenable for targeted therapeutics based on the absence or presence of specific genomic aberrations, this has increased the requirement for genomic profiling of tumor specimens.[1] Next-generation sequencing (NGS) allows for genomic characterization in a sensitive manner.[2] Although whole genome or exome sequencing both provide extensive genomic information, targeted gene panels are currently best fit for tumor profiling in a routine clinical context. It

Supported by Maag Lever Darm Stichting (MLDS) research grant FP13-9.
Disclosures: None declared.

http://dx.doi.org/10.1016/j.jmoldx.2016.06.010

matches best with the billable costs, short turn-around time (TAT), and requirement for reliable detection of variants compatible with routinely obtained material [formalin-fixed, paraffin-embedded (FFPE) tissue specimens, fine-needle aspirates, and cytologic smears].[3–5] This requires sufficient read depth to detect or exclude low-frequent variants that might be present due to tumor heterogeneity or a low tumor load in the tissue specimen and a robust and reliable bio-informatics pipeline. PCR-based amplification is commonly used to generate such targeted sequencing libraries for NGS.[6,7] Multiple genomic regions can be amplified simultaneously (multiplex PCR) for analysis of multiple genes from limited tissue material with low-quality genomic DNA (gDNA), such as the FFPE samples routinely used in molecular diagnostics. However, as a consequence of this amplification, true library complexity cannot be determined, because PCR duplicate reads cannot be distinguished from independent reads originating from separate original template molecules. This could result in overestimation of the actual number of DNA molecules analyzed, risking false-negative calls, which is crucial in the context of poor-quality samples with a small amount of amplifiable DNA. Single-molecule tagging (SMT) has been developed to overcome this issue by marking PCR duplicates originating from the same template molecule, which allows both a genuine analysis of library complexity and the possibility to combine multiple sequencing reads from PCR duplicates to generate a single consensus read.[8–10] The latter also allows filtering for errors originating during library amplification and sequencing that might result in false-positive calls due to jackpotting events.

We sought to develop an NGS-based targeted approach in a routine diagnostics setting for reliable detection of clinically relevant variants in tissue samples from FFPE specimens, in which multiplex analysis could be combined with SMT technology. Recently, multiplex analysis and SMT technology have been combined in single-molecule molecular inversion probes (smMIPs) to detect low-frequent variants in FFPE-derived DNA isolates in a simple, scalable, and relatively cost-effective manner.[11] In addition, the strand-specific nature of amplification by smMIPs can aid to distinguish genuine C:G>T:A mutations from those induced by cytosine deamination, a common artifact in DNA recovered from FFPE-fixed material.[12] Here, we describe the development, validation, and implementation of a single comprehensive smMIP-based Cancer Hotspot Panel (CHP) for mutation detection in clinically relevant genes.

Materials and Methods

Sample Preparation

Clinical specimens (generally from sections 3 × 20 μm) were digested at 56°C for at least 1 hour in the presence of TET-lysis buffer (10 mmol/L Tris/HCl pH8.5, 1 mmol/L EDTA pH8.0, 0.01% Tween-20) with 5% Chelex-100 (143 to 2832; Bio-Rad, Hercules, CA), 15 μg/mL GlycoBlue (AM9516; Thermo Fisher, Waltham, MA), and 400 μg proteinase K (19133; Qiagen, Valencia, CA), followed by inactivation at 95°C for 10 minutes. The supernatant was transferred after centrifugation and to reduce the total volume for the robotized protocol, cooled on ice and precipitated in the presence of 70% EtOH and 1/10 volume 3M NaAc (pH 5.2). Pellets were washed with cold 70% EtOH and dissolved in 80 μL Tris-EDTA, and DNA concentration was determined using the Qubit Broad Range Kit (Q32853; Thermo Fisher). Control DNA isolated from peripheral blood leukocytes was sonicated using a Covaris with a standard protocol to obtain 200-bp fragments and analyzed on a 1% agarose gel with 100-bp size ladder (15628-050; Invitrogen, Carlsbad, CA). The control NGS sample was obtained from Horizon Discovery (Waterbeach, UK; HD701).

Preparation of the smMIP-Pool for Targeted Enrichment

MIPs were designed using the procedure described[13] for all hotspots (Table 1), in a tiling manner preferentially covering all hotspots with two independent smMIPs targeting both DNA strands. The sum of the targeting arms is 40 bp (extension plus ligation probe arms) and the gap-fill length was set to 112 bp. The targeting arms were joined by a common backbone sequence and a stretch of 8 × N nucleotides was inserted between the backbone and ligation probe sequence (Supplemental Table S1). In case it was unavoidable to design smMIPs without common single nucleotide polymorphism variants in the respective capture arms, smMIPs were designed recognizing both alleles. Aliquots of each oligonucleotide smMIP probe (produced by Integrated DNA Technologies, Leuven, Belgium) were mixed in an equimolar or corrected fashion (Supplemental Table S1) to form the CHP smMIP-pool. The smMIP pool was phosphorylated with 1 μL of T4 polynucleotide kinase (M0201; New England Biolabs, Ipswich, MA) per 25 μL of 100 μmol/L smMIPs and ATP-containing T4 DNA ligase buffer (B0202; New England Biolabs). The molecular ratio between gDNA and smMIPs was set to 1:3200 for every individual smMIP (and is thus independent of pool size), and the required quantity of the smMIP pool was determined for a standard input of 100 ng gDNA.

Library Preparation

In manual experiments, a total of 100 ng of genomic DNA was used as input in a 20-μL volume, unless otherwise specified, with a total capture volume of 25 μL, including the (diluted) phosphorylated smMIP pool, 1 unit of Ampligase DNA ligase (A0110K; EpiBio, Madison, WI) with Ampligase Buffer (A1905B, DNA ligase buffer), 3.2 units of Hemo Klentaq (M0332; New England Biolabs), and 8 μmol of dNTPs (28-4065-20/-12/-22/-32; GE Healthcare, Little Chalfont, UK). After denaturation (95°C for 10