The Journal of Molecular Diagnostics, Vol. 18, No. 3, May 2016

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Plasmid-Based Materials as Multiplex Quality Controls and Calibrators for Clinical Next-Generation Sequencing Assays

David J. Sims,* Robin D. Harrington,* Eric C. Polley,† Thomas D. Forbes,* Michele G. Mehaffey,* Paul M. McGregor, III,* Corinne E. Camalier,* Kneshay N. Harper,* Courtney H. Bouk,* Biswajit Das,* Barbara A. Conley,† James H. Doroshow,† P. Mickey Williams,* and Chih-Jian Lih*

From the Molecular Characterization and Clinical Assay Development Laboratory, Frederick National Laboratory for Cancer Research, Frederick; and the Division of Cancer Treatment and Diagnosis,† National Cancer Institute, Bethesda, Maryland*

Accepted for publication November 20, 2015.

Address correspondence to Chih-Jian Lih, Ph.D., Frederick National Laboratory for Cancer Research, Bldg. 320, Room 5, 1050 Boyles St., Frederick, MD 21702. E-mail: jason.lih@nih.gov.

Although next-generation sequencing technologies have been widely adapted for clinical diagnostic applications, an urgent need exists for multianalyte calibrator materials and controls to evaluate the performance of these assays. Control materials will also play a major role in the assessment, development, and selection of appropriate alignment and variant calling pipelines. We report an approach to provide effective multianalyte controls for next-generation sequencing assays, referred to as the control plasmid spiked-in genome (CPSG). Control plasmids that contain approximately 1000 bases of human genomic sequence with a specific mutation of interest positioned near the middle of the insert and a nearby 6-bp molecular barcode were synthesized, linearized, quantitated, and spiked into genomic DNA derived from formalin-fixed, paraffin-embedded—prepared hapmap cell lines at defined copy number ratios. Serial titration experiments demonstrated the CPSGs performed with similar efficiency of variant detection as formalin-fixed, paraffin-embedded cell line genomic DNA. Repetitive analyses of one lot of CPSGs 90 times during 18 months revealed that the reagents were stable with consistent detection of each of the plasmids at similar variant allele frequencies. CPSGs are designed to work across most next-generation sequencing methods, platforms, and data analysis pipelines. CPSGs are robust controls and can be used to evaluate the performance of different next-generation sequencing diagnostic assays, assess data analysis pipelines, and ensure robust assay performance metrics. *(J Mol Diagn 2016, 18: 336—349; http://dx.doi.org/10.1016/j.jmoldx.2015.11.008)*

Next-generation sequencing (NGS) technology is having major effects on biomedical research. Decreasing costs and increasing data generation are driving rapid uptake of this method. Clinical applications have quickly followed.[1,2] NGS technology is currently under evaluation for guiding cancer patient treatment selection.[3,4] However, there is uncertainty that there is sufficient interlaboratory concordance for meaningful clinical use. The rapid proliferation of different sequencing methods, platforms, and data analysis tools has resulted in a high discordance of mutations reported from different clinical NGS assays.[5,6] Reference and control materials that contain known analytes (variants) at known allele fraction [variant allele frequency (VAF)] in a form comparable to clinical specimens are essential for comparing and monitoring the assay performance and will be valuable in the study of cross-platform comparisons and identifying weaknesses in informatics pipelines (ie, alignment and variant calling). However, unlike most conventional assays (eg, Sanger sequencing and PCR-based methods) that typically detect single or only a few analytes,

Supported by National Cancer Institute, NIH, grants HHSN261200800001E and NO1-CO-2008-00001.

This work does not express or represent the opinion of the National Cancer Institute, National Institutes of Health, or Department of Health and Human Service.

Disclosures: None declared.

Current address of M.G.M., Department of Pediatrics, University of Washington, Seattle, WA.

Published by Elsevier Inc. on behalf of the American Society for Investigative Pathology and the Association for Molecular Pathology.
http://dx.doi.org/10.1016/j.jmoldx.2015.11.008

an NGS assay usually measures hundreds to thousands of genomic loci. Currently, there is no standardized set of clinically relevant materials useful as controls or calibrators to standardize the assessment of NGS data across platforms, assays, and informatics pipelines. Genome in a Bottle, a public consortium led by the National Institute of Standards and Technology, has released a reference genome and will soon release several other genomes.[7] These are valuable resources but do not directly address the need for clinically relevant controls and calibrators. Therefore, there is an urgent need to implement highly multiplexed materials as calibrators and controls for the clinical use of NGS assays.[5,6,8]

One approach to NGS calibrators and controls relies on the use of cell line genomic DNA. A mixture of variant types and VAF can be manufactured by combining genomes at defined molar ratios.[9] This approach is limited by the number of genomes that can be mixed while maintaining an adequate VAF and by the number of different mutations that can be introduced into a single cell line.

Another approach is the use of synthetic nucleic acid molecules, such as long oligonucleotides as used in the SNaPshot assay[10] and *in vitro* transcribed RNA molecules from the External RNA Control Consortium (ERCC) used in gene expression and RNAseq assays.[11] In taking the first step toward building highly multiplex control materials, we report the development and characterization of a control plasmid–based multianalyte calibrator and control material for NGS assays, termed the control plasmid spiked-in genome (CPSG). We found that these materials are scalable in their ability to incorporate many different variants with different allele frequencies in a complex mixture, are easy to design and manufacture, are distinguishable from a clinical specimen, and are detectable by various genomic assays. Our results indicate that CPSGs can serve as routine assay controls to monitor performance of NGS assays and standards for cross-site and cross-platform comparison studies and as valuable tools for the evaluation, development, and testing of new informatics pipelines. Such an approach was previously accepted by the US Food and Drug Administration as an effective method of validating the detection of rare germline variants with an NGS platform in a submission of 510 (k) premarket notification (Food and Drug Administration, *http://www.accessdata.fda.gov/cdrh_docs/pdf13/K132750.pdf*, last accessed November 20, 2015) by Illumina (Illumina MiSeqDx Cystic Fibrosis Clinical Sequencing Assay; Illumina Inc., San Diego, CA). Importantly, we also found that the efficiency of variant detection in CPSG samples is similar to that of formalin-fixed, paraffin-embedded (FFPE) genomic DNA samples.

Materials and Methods

Design and Construction of Control Plasmids

To evaluate the performance of various NGS assays on different types of mutations, a panel of 69 control plasmids was designed and constructed and a subset of them used for this study. This panel of 69 control plasmids contains 38 single-nucleotide variants (SNVs), nine SNVs at a homopolymeric region (HP; >3 identical bases in a row), 12 insertion/deletions (indels), five indels at HP, and five large indels (gap size >4 bp). Mutations of interest (MOIs) in these control plasmids were selected because of their known clinical actionable value and high recurrent frequency in the Catalogue of Somatic Mutations in Cancer database or because they represent rare mutation types. For each MOI, an approximate 1000-bp region flanking (approximately 500 bp upstream and approximately 500 bp downstream) the MOI was synthetically generated (DNA 2.0, Menlo Park, CA). In addition, a 6-bp insert sequence (ACATCG), which functions as a molecular barcode, was placed 5 to 20 bp away from the

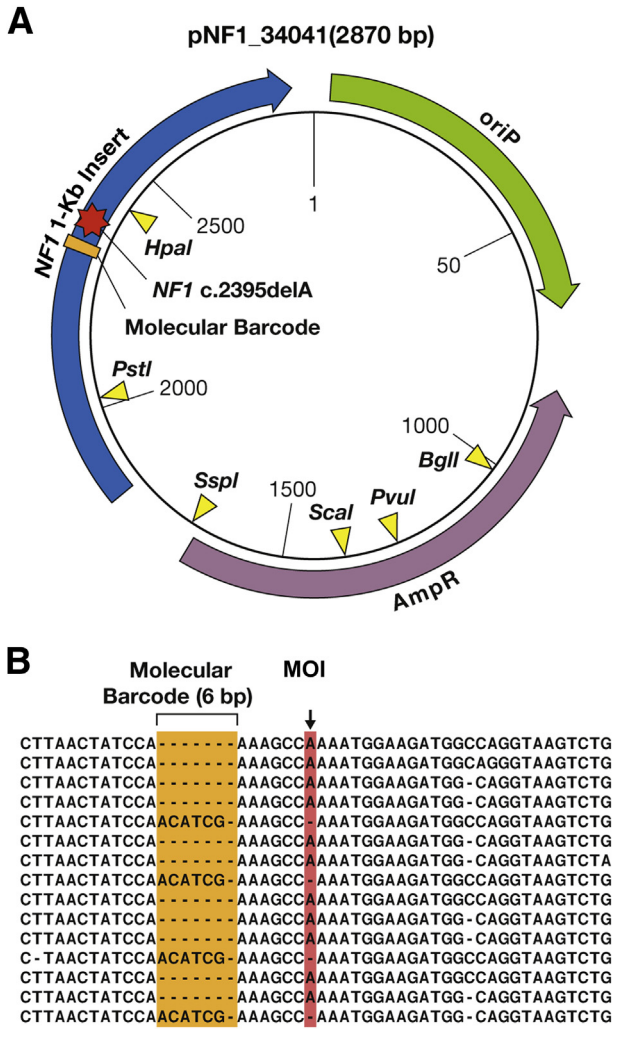

Figure 1 Design of control plasmids. **A:** Map of a representative control plasmid, pNF1_34041. Each control plasmid was constructed by inserting approximately 1000 bp of genomic DNA (blue box) spanning the mutation of interest (MOI) (**red star**). A 6-bp (ACATCG) molecular barcode (**orange rectangle** in **B**) was inserted near the MOI to track variant reads. Single-cut restriction sites are indicated by yellow triangles. **B:** Coordination of the molecular barcode with the MOI. A subset of sequencing reads from MOIs [an A deletion (**red box**)] and 6-bp molecular barcode confirms the mutation is plasmid borne.