Clinical Chemistry 60:9
1192–1199 (2014)

Microsatellite Instability Detection by Next Generation Sequencing

Stephen J. Salipante,[1] Sheena M. Scroggins,[1] Heather L. Hampel,[2] Emily H. Turner,[1] and Colin C. Pritchard[1]*

BACKGROUND: Microsatellite instability (MSI) is a useful phenotype in cancer diagnosis and prognosis. Nevertheless, methods to detect MSI status from next generation DNA sequencing (NGS) data are underdeveloped.

METHODS: We developed an approach to detect the MSI phenotype using NGS (mSINGS). The method was used to evaluate mononucleotide microsatellite loci that were incidentally sequenced after targeted gene enrichment and could be applied to gene or exome capture panels designed for other purposes. For each microsatellite locus, the number of differently sized repeats in experimental samples were quantified and compared to a population of normal controls. Loci were considered unstable if the experimental number of repeats was statistically greater than in the control population. MSI status was determined by the fraction of unstable microsatellite loci.

RESULTS: We examined data from 324 samples generated using targeted gene capture assays of 3 different sizes, ranging from a 0.85-Mb to a 44-Mb exome design and incorporating from 15 to 2957 microsatellite markers. When we compared mSING results to MSI-PCR as a gold standard for 108 cases, we found the approach to be both diagnostically sensitive (range of 96.4% to 100% across 3 panels) and specific (range of 97.2% to 100%) for determining MSI status. The fraction of unstable microsatellite markers calculated from sequencing data correlated with the number of unstable loci detected by conventional MSI-PCR testing.

CONCLUSIONS: NGS data can enable highly accurate detection of MSI, even from limited capture designs. This novel approach offers several advantages over existing PCR-based methods.

Microsatellite instability (MSI)[3] is a molecular phenotype marked by the spontaneous acquisition or loss of nucleotides from within repetitive microsatellite tracts, which results in the production of novel microsatellite alleles *(1, 2)*. The underlying mechanism of MSI is dysregulation of the mismatch repair (MMR) system, which limits cells' ability to correct spontaneous, length-altering somatic mutations that occur with high frequency in microsatellites *(3)*. MSI tumors may result from inactivating germline mutations in one or more genes, including MutL homolog 1, colon cancer, nonpolyposis type 2 (*E. coli*) (*MLH1*),[4] DNA mismatch repair protein Msh2 (*MSH2*), mutS homolog 6 (*MSH6*), mismatch repair endonuclease PMS2 (*PMS2*), and epithelial cell adhesion molecule (*EPCAM*), such as occurs in patients with Lynch syndrome, for whom more than 90% of colon cancers test MSI positive *(1, 2)*. MSI also occurs sporadically in several cancer types, including colorectal, endometrial, ovarian, and gastric cancers. In contrast to Lynch syndrome, sporadic MSI is often due to somatic promoter hypermethylation of *MLH1* in the absence of gene sequence mutations *(1, 4)*.

Diagnosis of MSI often has important clinical implications *(5–7)*, informing therapeutic choices, cancer prognosis, and familial cancer risk appraisal *(1, 2, 8–10)*. PCR detection of instability at informative microsatellite markers (MSI-PCR) is the chief DNA-based method in current clinical use *(6)*. Select microsatellites are PCR amplified using fluorescently labeled primers, and fragment length polymorphisms are identified through capillary electrophoresis. Current guidelines for MSI-PCR recommend the use of relatively small, standardized panels of microsatellites that include highly unstable mononucleotide repeat loci *(11)*. Such panels have been evaluated both for relative instability in an MSI background and genetic mono-

[1] Department of Laboratory Medicine, University of Washington, Seattle WA; [2] Department of Internal Medicine, Division of Human Genetics, The Ohio State University, Columbus, OH.
* Address correspondence to this author at: University of Washington, 1959 NE Pacific St., Box 357110, Seattle, WA, 98195. Fax 206-598-6189; e-mail cpritch@uw.edu.
Received February 21, 2014; accepted May 15, 2014.
Previously published online at DOI: 10.1373/clinchem.2014.223677

[3] Nonstandard abbreviations: MSI, microsatellite instability; MMR, mismatch repair; NGS, next-generation DNA sequencing; mSINGS, determination of MSI by NGS; TCGA, the Cancer Genome Atlas Network.

[4] Human genes: *MLH1,* MutL homolog 1, colon cancer, nonpolyposis type 2 (*E. coli*); *MSH2,* DNA mismatch repair protein Msh2; *MSH6,* mutS homolog 6; *PMS2,* Mismatch repair endonuclease PMS2; *EPCAM,* epithelial cell adhesion molecule.

morphism at the population level *(1, 12)*, which simplifies their diagnostic interpretation.

Studies of cancer genetics are increasingly making use of massively parallel or next generation DNA sequencing (NGS) technologies, both through use of whole-genome or whole-exome sequencing *(13)* and by more targeted sequencing assays as clinical diagnostic methods *(14–16)*. NGS provides an unprecedented economy of scale, allowing dozens to hundreds of genes to be sequenced simultaneously for each patient and with higher sensitivity for low-prevalence mutations *(16, 17)*. Yet, because the majority of sporadic MSI-positive tumors often result from epigenetic changes rather than coding mutations in MMR genes *(1, 4)*, even fully sequencing all MMR pathway genes by NGS *(18)* will not provide sufficient data to reliably infer MSI status in a tumor. Alternatively, determination of epigenetic methylation patterns by NGS requires complex library preparation techniques *(19)*, which also prevent the reliable detection of mutations.

Here we describe an approach for determination of MSI by NGS (mSINGS) based on microsatellite markers which are incidentally included in targeted gene capture sequencing data. Among other benefits, this method allows MSI status to be reliably deduced at the same time that other genes of interest are sequenced, without the need for dedicated inclusion of specific markers.

Materials and Methods

SAMPLES AND SEQUENCING DATA

We examined NGS data obtained from 324 total samples prepared by 3 different methods. (*a*) Exome-sequencing data from 10 known MSI-negative, 12 known MSI-positive, and 4 known MSI-low colorectal cancers, generated by the Cancer Genome Atlas Network (TCGA) *(13)*. Per TGCA protocols, samples were fresh-frozen material with ≥60% tumor purity *(13)*. (*b*) Targeted gene sequencing data from the ColoSeq assay (University of Washington, Seattle, WA) *(20)* for 103 samples, including colorectal, endometrial, ovarian, breast, and prostate tumors. Samples were formalin fixed paraffin embedded, with tumor purity ≥20%. (*c*) Targeted gene sequencing data from the UW-OncoPlex assay (University of Washington) *(14)* for 195 tumor samples, including colon, endometrial, lung, breast, ovarian, melanoma, and additional tumor samples. Samples were formalin fixed paraffin embedded, with tumor purity ≥20% based on review of hematoxylin and eosin–stained slides. Data were from samples tested between November 2011 and December 2013. Clinical samples and data were obtained in accordance with the declaration of Helsinki and the ethics guidelines of the Human Subjects Division of the University of Washington and the Ohio State University.

The 3 capture assays varied substantially in design and the amount of genomic sequence captured. Exome data generated by TCGA comprised approximately 44 Mb of exonic sequence data from roughly 30 000 genes. Targeted gene capture designs for the 2 clinical sequencing assays, ColoSeq *(18)* and UW-OncoPlex *(14)*, captured approximately 1.4 Mb and 0.85 Mb of sequence data, respectively. The ColoSeq capture provided full sequence coverage of 50 genes, including exons, most introns, 5′ and 3′ untranslated regions, and some flanking intergenic sequences. The UW-OncoPlex included exonic sequences from 194 genes and selected introns of particular genes implicated in structural rearrangements.

MSI-PCR TESTING

MSI-PCR testing was performed for a subset of samples by the University of Washington Clinical Molecular Genetics laboratory using the Promega MSI analysis kit (Promega). Unless specified otherwise, samples demonstrating instability of 2 or more of the 5 mononucleotide markers included in this panel were considered MSI positive, and others were considered MSI negative. For the 2 clinical assays ColoSeq and UW-OncoPlex, MSI-PCR was performed for a subset of samples as the gold standard method for determining MSI status (n = 64 for ColoSeq and n = 18 for UW-OncoPlex). TCGA exome data had previously undergone diagnostic MSI-PCR (n = 26). Analysts (n = 2 MD, PhD clinical laboratory directors) were kept blinded to MSI status if this information was available in advance.

DATA PREPROCESSING

Initial read mapping against the human reference genome (hg19/GRCh37) and alignment processing were performed using BWA version 0.6.1-r104 *(21)* and SAMtools version 0.1.18 *(22)*. For TCGA data, pre-aligned reads were obtained from the NCI Cancer Genomics Hub repository.

Sample-level, fully local indel realignment was then performed using GATK version 2.4 *(23)*. Duplicate reads were removed using Picard version 1.72 (http://picard.sourceforge.net). Quality score recalibration was then performed using GATK to generate a final realigned and recalibrated alignment, which was used for subsequent analyses.

Indel calling was performed through VarScan version 2.2.8 *(24)*, with the minimum variant frequency set to 0.01 reads and the minimum number of variant reads set to 4. Read counts for each indel of a unique length were quantified using VarScan with a minimum

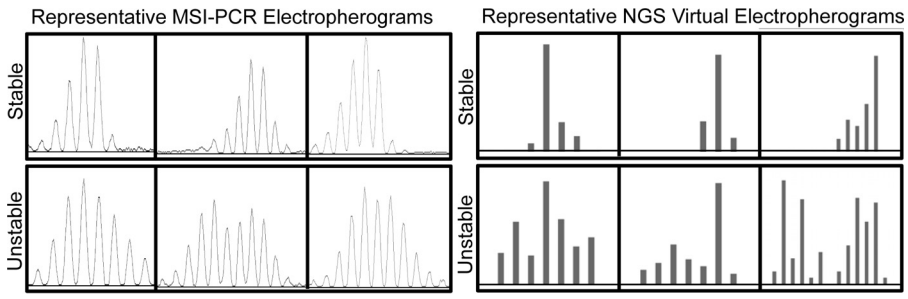

Fig. 1. Detection of microsatellite instability by MSI-PCR and NGS.

Representative capillary electrophoresis results from MSI-PCR (left) and virtual electropherograms of NGS data (right). The length (*x* axis) and relative abundance (*y* axis) of variant repeats are plotted. Columns represent individual loci and show paired data for MSI-stable (top) and -unstable (bottom) samples. Different microsatellite loci are shown for MSI-PCR and NGS and are not directly comparable.

base quality of 10 and all other parameters at the default.

MICROSATELLITE LOCI IDENTIFICATION

To enable evaluation of MSI status, we first identified genomic sites with evidence of length instability on sequencing. Suitable microsatellite loci were determined independently for each of the 3 panels. We examined variant call data of a subset of known MSI-positive tumor samples (n = 3) to identify sites with 3 or more different length polymorphisms called. This empiric approach did not require a minimum microsatellite tract length, base composition, or repeat unit size (such as mononucleotide, dinucleotide, or trinucleotide). Regardless, for each capture design, the vast majority (>99%) of sites corresponded to mononucleotide tracts (monotonous runs of A/T or C/G). Because mononucleotide repeats are believed to be most sensitive and specific for detecting MSI *(25)*, we limited our final microsatellite panel to that class of markers. Identified sites were extended out by 5–35 bp of flanking sequence to define the final interrogated range for each locus. For the smaller targeted assays, sites were identified by manual review of primary variant calls, whereas for exome data this process was performed using a custom Perl script (available on request). We identified 2957 suitable microsatellite marker loci in exome data, 146 loci in ColoSeq, and 15 loci for UW-OncoPlex. Genomic coordinates (hg19/GRCh37) of microsatellite loci are given in Tables 1–3 in the Data Supplement that accompanies the online version of this report at http://www.clinchem.org/content/vol60/issue9.

DETERMINING CHARACTERISTICS OF LOCI IN MSI-STABLE SAMPLES

Amplifying microsatellite loci by PCR generates a distribution of alternate fragments (also referred to as "stutter" artifact) that results from slippage during PCR amplification (Fig. 1) *(26, 27)*. We therefore first calculated descriptive statistics about the number of slippage fragments observed at each locus across a population of MSI-negative control samples to establish baseline reference values.

For each control sample we evaluated the number of repeats of different lengths present within each of the identified microsatellite markers. We calculated the baseline reference value as the mean number of unique repeat lengths at each mononucleotide tract across a population of MSI-negative samples. For exome and ColoSeq data, we generated a baseline from nontumor peripheral blood DNA samples (n = 11 and n = 47, respectively). For UW-OncoPlex we established a baseline from a subset of randomly selected MSI-negative tumors (n = 10). A baseline was established through the use of custom scripts as follows. (*a*) For each locus, we required a minimum read depth of 30 or more reads to be included in the baseline calculation. (*b*) The number of alleles detectable at a site should be proportional to read depth at a site, with greater numbers of reads resulting in a greater ability to discriminate low-prevalence alleles. To normalize the number of alleles observed in a sample with respect to the read depth, the number of reads from alleles of each observed length compared to the reference genome (i.e., −2, −1, +1, and +2) were expressed as a percentage of the number of reads counted for the most frequently occurring allele. (*c*) Alleles with <5% of the reads counted for the most frequently observed allele were excluded. Tumor instability could potentially be reflected by low-prevalence alleles present at <5% abundance. However, because the efficiency with which nonprevalent alleles are detected is a function of the read depth at a locus, it was necessary to assign a 5% cutoff to allow comparison among samples with disparate amounts of

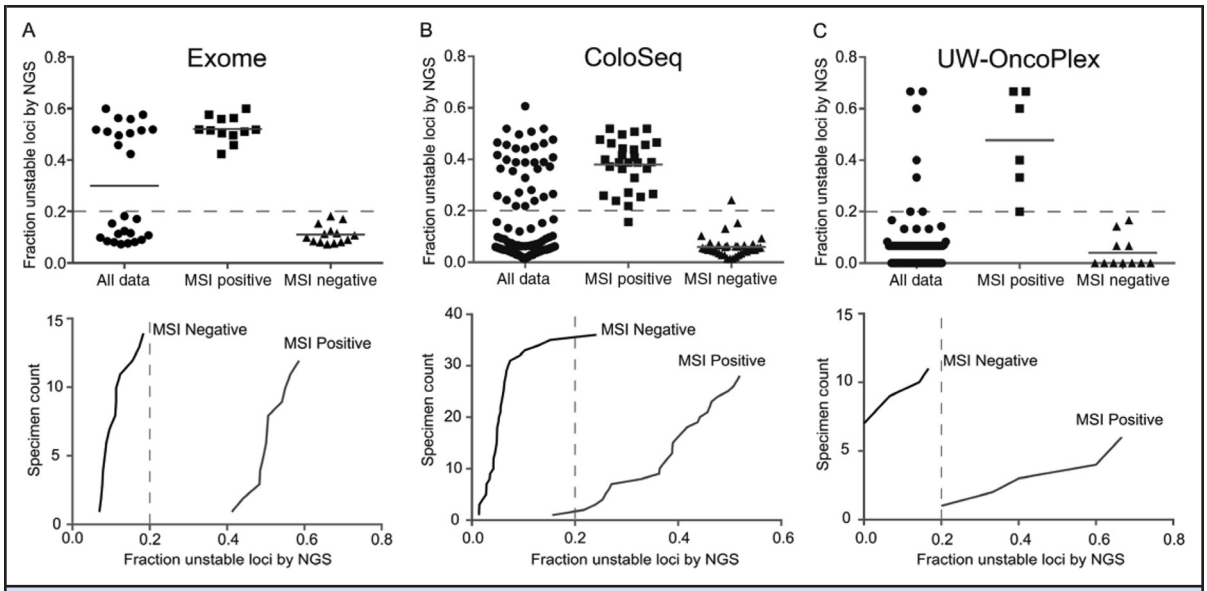

Fig. 2. Fraction of unstable microsatellite loci identified by NGS predicts MSI status.

The fraction of unstable microsatellite loci (top) and cumulative distribution of that fraction (bottom) are shown for the targeted gene capture designs exome capture (A), ColoSeq (B), and UW-OncoPlex (C). Results are stratified by conventional MSI-PCR status, with the mean fraction of unstable loci indicated by a solid horizontal line. The threshold used for interpreting MSI status is indicated by a dashed line.

sequence coverage. (*d*) For each locus, we calculated the mean and SD of the number of alleles. In general, most samples demonstrated sufficient read coverage to contribute to the calculation of summary statistics at all loci. However, rare loci for which fewer than 3 samples could be used to generate summary statistics were excluded from the final panel.

EXPERIMENTAL DETERMINATION OF MSI STATUS

We compared experimental results against baseline reference values at each locus to assess the instability of microsatellite loci. Data were processed in a similar fashion as in establishing normal control baselines. For each sample, microsatellite loci covered by a read depth of <30 were not considered and instead were reported as missing information. At each locus passing this initial QC check, the number of reads from each observed allele was expressed as a percentage of the number of reads counted for the most frequently occurring allele. After this normalization, the total number of repeats of different lengths with a read count exceeding 5% or more of that seen for the most frequently observed allele was tallied. This tally was compared against the baseline for the same microsatellite locus. If the tally of alleles counted exceeded [mean number of alleles + ($3 \times$ SD)] the MSI stable reference value, the locus was scored as unstable, and if the tally did not exceed this value, it was scored as stable. This metric provides a statistical framework for evaluating the stability or instability of any particular marker. Finally, the fraction of unstable loci out of the total number of loci analyzed was calculated for each experimental sample. In comparing microsatellites, we observed considerable variability in the frequency that particular loci were classified as unstable among samples (see online Supplemental Tables 1–3).

SOURCE CODE

Source code for mSINGs analysis is available through https://bitbucket.org/uwlabmed/msings.

Results

DETERMINATION OF MSI BY NGS DATA

We performed mSINGS on 324 samples sequenced using 3 capture designs of different sizes (Exome, ColoSeq, and UW-OncoPlex). For each assay, samples could be divided into 2 populations (Fig. 2) characterized by a high or low fraction of microsatellite unstable loci. Based on this qualitative separation and existing guidelines for defining MSI positivity *(12)*, we selected a cutoff fraction of 0.2 (20%) unstable loci for an MSI-positive result for each assay.

MSI status by mSINGS had high diagnostic sensitivity and specificity for each capture design (Table 1). The MSI phenotype was correctly inferred for all sam-

Table 1. Diagnostic sensitivity and specificity of NGS panels for MSI detection using mSINGS.

Capture panel	Number of loci	Sensitivity,[a] no. positive by mSINGS/ no. positive by MSI-PCR (% [95% CI[b]])	Specificity,[a] no. negative by mSINGS/ no. negative by MSI-PCR (% [95% CI])
Exome	2957	12/12 (100 [73.5–100])	14/14 (100 [76.8–100])
ColoSeq	146	27/28 (96.4 [85.5–99.9])	35/36 (97.2 [81.7–99.9])
UW-OncoPlex	15	6/6 (100 [54.1–100])	12/12 (100 [73.5–100])
Combined	15–2957	45/46 (97.8 [88.5–99.9])	61/62 (98.3 [91.3–99.9])

[a] Calculations are based on assumption of MSI-PCR as a gold standard. Only the subset of 108 out of 324 total samples analyzed that had gold standard MSI-PCR data available are included.
[b] All CIs calculated by the Clopper-Pearson (exact) method.

ples examined by both exome capture and UW-OncoPlex, corresponding to a sensitivity and specificity of 100%. The MSI phenotypes of 2 samples tested by MSI-PCR were discordant with ColoSeq NGS data. One sample was positive by PCR-based MSI testing (2 of 5 loci unstable) but fell below our cutoff for MSI positivity by NGS, with a fraction of 0.16 mutated alleles. The other was negative by conventional MSI-PCR testing (0 out of 5 unstable loci) but fell slightly above our cutoff for declaring a sample to be MSI positive, with a fraction of 0.24 mutated alleles. The assay consequently demonstrates a diagnostic sensitivity of 96.4% and a specificity of 97.2% for inferring MSI status with this capture design (Table 1).

The fraction of unstable loci detected by mSINGS was strongly correlated to MSI status by conventional MSI-PCR (Fig. 2). For exome data, MSI-positive tumors had a mean fraction of 0.52 unstable loci (12 samples; range, 0.42–0.60; SD, 0.05), whereas MSI-negative samples displayed a significantly lower mean fraction of 0.11 unstable loci (14 samples; range, 0.07–0.18; SD, 0.04; $P = 5.34 \times 10^{-16}$, 2-tailed *t*-test). In the ColoSeq capture design, MSI-positive tumors had a mean fraction of 0.38 unstable loci (28 samples; range, 0.16–0.52; SD, 0.97), and MSI-negative samples demonstrated a significantly ($P = 1.08 \times 10^{-17}$, 2-tailed *t*-test) lower mean fraction of 0.06 unstable loci (36 samples; range, 0.01–0.24; SD, 0.42). Similarly, MSI-PCR–positive samples tested by UW-OncoPlex had a mean fraction of 0.44 (6 samples; range, 0.20–0.67; SD, 0.17) unstable loci each, whereas MSI-PCR–negative samples had a significantly lower mean fraction of unstable loci at 0.04 (12 samples; range, 0–0.17; SD, 0.06; $P = 0.002$, 2-tailed *t*-test).

Four tumors sequenced by exome and 1 tumor sequenced by UW-OncoPlex carried an MSI-low designation, conventionally used to define samples with some number of unstable loci insufficient to qualify for definite MSI positivity *(7)*. For our primary analysis these samples were considered MSI negative *(6)*. All MSI-low samples fell below the cutoff for MSI positivity by NGS. MSI-low samples did not differ significantly from confirmed MSI-negative samples: exome-sequenced MSI-low samples demonstrated a mean of 0.13 unstable sites (range, 0.08–0.18; SD, 0.05; $P = 0.38$, 2-tailed *t*-test), and the single MSI-low UW-OncoPlex sample had 0 unstable sites (2-tailed Z-test $= -0.76$).

QUANTITATIVE CORRELATION OF MSI-PCR AND mSINGS

We evaluated whether there was a correlation between the fraction of mutated loci detected by conventional MSI-PCR testing and the fraction of mutated loci identified by mSINGS. Detailed MSI-PCR results were available for only ColoSeq and UW-OncoPlex samples, so analysis was limited to those panels. We found a high degree of correlation between the fraction of unstable loci detected by conventional testing and by mSINGS (R^2 values of 0.86 and 0.94, respectively) (Fig. 3).

Discussion

We have developed a strategy for inferring a tumor's MSI phenotype through NGS of microsatellite loci that are incidentally captured during targeted sequencing of gene panels, which we have termed mSINGS. We applied the approach to 3 capture designs ranging from 0.85 to 44 Mb of genomic sequences and found in each case that NGS could be used to infer MSI status with high sensitivity and specificity (Fig. 2 and Table 1).

The criteria we employed for determining a sample's MSI phenotype are related to those currently used in interpreting MSI-PCR assays, in that we compare the number of signals reflecting products of different lengths at an individual marker against those from a healthy sample to assess potential instability (Fig. 1). However, to make the approach both as standardized and as broadly applicable as possible, our method compares each marker to a population of MSI-negative samples, rather than to a single-sample negative control. This allows for loci to be assessed for statistically significant deviations from a control distribution, pro-

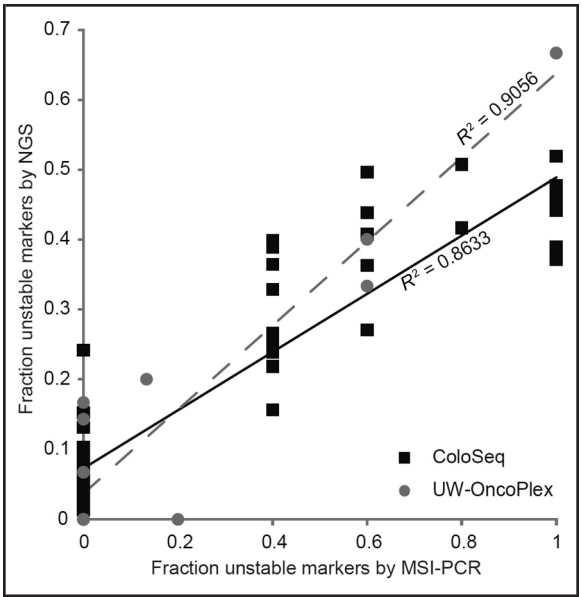

Fig. 3. Quantitative correlation of MSI-PCR and NGS results.

The fraction of unstable microsatellite markers as determined by NGS is plotted with relation to the fraction of unstable markers detected by conventional MSI-PCR. Samples analyzed by ColoSeq and UW-OncoPlex assays are displayed separately, with linear trend lines (dashed grey line and solid black line, respectively) and corresponding R^2 values.

viding a well-defined and reproducible standard for defining a locus as unstable. Even when using panels of microsatellite markers that have not been extensively vetted in the same fashion as loci used for MSI-PCR assays *(1, 12)*, the variation due to germline microsatellite locus polymorphisms and other potential sources of underlying genomic variation do not appear limiting. Comparison of NGS data against summary statistics calculated from a population of MSI-negative samples has proven to be a tractable strategy for identifying MSI and may eliminate the need for patient-matched nontumor controls.

A robust approach for permitting detection of MSI phenotypes using NGS may offer several additional advantages over conventional PCR testing. First, use of NGS data for MSI detection could eliminate the need for separate, dedicated testing for MSI status, resolving this clinically important tumor phenotype during the course of obtaining more comprehensive genetic information. This approach would reduce healthcare costs while also conserving sometimes limited sample materials. Microsatellite markers are abundant in the genome *(6, 27)*, such that they will be recovered in sufficient numbers by most targeted gene panels. In the 3 targeted assays we examined, we identified numbers of microsatellite loci sufficient to enable accurate MSI analysis without special considerations having been previously taken into account during capture design. However, it appears that some loci are more informative than others (see online Supplemental Tables 1–3), suggesting that mSINGS performance is not strictly a function of the number of microsatellites that can be examined.

A second advantage of an NGS approach is that its scalability allows a far greater number of microsatellite markers to be examined than is practical or cost-effective by PCR-based testing. Each of the 3 assays considered in this study already contains more markers than present in the standard MSI-PCR panel. Given the relatively small size of microsatellite tracts, however, loci could be specifically targeted during design of gene capture assays, theoretically increasing the number of informative targets for MSI analysis with a minimal impact on overall capture or sequencing efficiency. We suggest that, when possible, targeted gene capture designs intended for mSINGS analysis include conventional MSI markers *(11)* to enable internal comparison with gold standard methods. Although PCR-based MSI testing of small panels of markers performs well *(11)*, examination of increasingly large numbers of informative markers is likely to enable greater sensitivity and specificity than existing MSI testing approaches.

The capacity of NGS to examine large numbers of loci also offers a means to quantify the degree of MSI displayed by a tumor to an extent not previously achievable. The importance of various degrees of MSI, if any, is currently unclear, especially in distinguishing MSI-low from MSI-negative tumors *(6, 7, 28)*, but this enhanced resolving power may yield additional diagnostic information with further study. Here, the small numbers of available MSI-low tumors were not distinguishable from MSI-negative samples by NGS, suggesting either that the MSI-low designation does not represent a distinct phenotype or that the MSI-low phenotype was not detectable using our approach. Larger, dedicated studies of MSI-low samples are required to draw firm conclusions, and similarly, to investigate whether MSI-positive tumors demonstrating a higher fraction of unstable loci are biologically or clinically different from MSI-positive tumors with a lower fraction of unstable sites. Exploring whether the pattern of unstable loci observed differs across different tumor types also deserves special attention and is warranted in future studies.

Because every DNA fragment is assayed independently during NGS, the technology provides accurate digital information about the numeric count and relative abundance of each alternative repeat length ob-

served at a locus. In contrast to PCR amplification and capillary electrophoresis, assays for MSI utilizing NGS can provide quantitative, rather than qualitative, information about the number and distribution of novel microsatellite repeat-length polymorphisms detected. In this work we employed quantitative statistics to define whether a locus was unstable, based on the mean number of unique repeat lengths observed across a nonmutated population. Defining whether a locus is unstable by PCR and capillary electrophoresis, which generates an analog signal, is more subjective and can be especially challenging if alleles are close together in size *(27)*. Use of NGS enables more principled characterization of markers as stable or unstable, which might help to standardize interpretation of results.

In summary, evaluation of the MSI phenotype by NGS is valuable not only as an ancillary analysis provided from large-scale sequencing assays, but also offers advantages for MSI diagnosis as a primary assay. As the cost of NGS technologies continues to drop, routine laboratory diagnosis of the MSI phenotype by this approach will become increasingly tractable and affordable. Such an approach could improve the quality of MSI diagnosis, standardize interpretation, and provide value-added information.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: H.L. Hampel, received honorarium for talk to Quest national sales force.
Research Funding: C.C. Pritchard, 2013 Young Investigator Award from the Prostate Cancer Foundation.
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: No sponsor was declared.

Acknowledgments: We thank Jonathan Tait and William Grady for help reviewing the manuscript. We thank Deborah Barden, Rachel Slusher, Laura Akagi, Youly Welt, Christina Smith, Karen Koehler, and Tatyana Marushchak for help with genomic DNA, library preparation and with MSI-PCR.

References

1. Murphy KM, Zhang S, Geiger T, Hafez MJ, Bacher J, Berg KD, Eshleman JR. Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers. J Mol Diagn 2006;8:305–11.
2. Zhang L. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part II. The utility of microsatellite instability testing. J Mol Diagn 2008;10:301–7.
3. Oki E, Oda S, Maehara Y, Sugimachi K. Mutated gene-specific phenotypes of dinucleotide repeat instability in human colorectal carcinoma cell lines deficient in DNA mismatch repair. Oncogene 1999;18:2143–7.
4. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: molecular basis of colorectal cancer. N Engl J Med 2009;361:2449–60.
5. Kim TM, Laird PW, Park PJ. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell 2013;155:858–68.
6. de la Chapelle A, Hampel H. Clinical relevance of microsatellite instability in colorectal cancer. J Clin Oncol 2010;28:3380–7.
7. Pawlik TM, Raut CP, Rodriguez-Bigas MA. Colorectal carcinogenesis: MSI-H versus MSI-L. Dis Markers 2004;20:199–206.
8. Sankila R, Aaltonen LA, Jarvinen HJ, Mecklin JP. Better survival rates in patients with MLH1-associated hereditary colorectal cancer. Gastroenterology 1996;110:682–7.
9. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med 2003;349:247–57.
10. Beamer LC, Grant ML, Espenschied CR, Blazer KR, Hampel HL, Weitzel JN, MacDonald DJ. Reflex immunohistochemistry and microsatellite instability testing of colorectal tumors for Lynch syndrome among US cancer programs and follow-up of abnormal results. J Clin Oncol 2012;30:1058–63.
11. Bacher JW, Flanagan LA, Smalley RL, Nassif NA, Burgart LJ, Halberg RB, et al. Development of a fluorescent multiplex assay for detection of MSI-High tumors. Dis Markers 2004;20:237–50.
12. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248–57.
13. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.
14. Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. J Mol Diagn 2014;16:56–67.
15. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer Discov 2012;2:82–93.
16. Harismendy O, Schwab RB, Bao L, Olson J, Rozenzhak S, Kotsopoulos SK, et al. Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. Genome Biol 2011;12:R124.
17. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 2013;23:843–54.
18. Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, Boerno ST, et al. Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. PloS One 2010;5:e15661.
19. Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, et al. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res 2007;67:8511–8.
20. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn 2012;14:357–66.
21. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589–95.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for

variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.
24. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. Varscan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.
25. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261–8.
26. Clarke LA, Rebelo CS, Goncalves J, Boavida MG, Jordan P. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol 2001;54:351–3.
27. Salipante SJ, Horwitz MS. Phylogenetic fate mapping. Proc Natl Acad Sci U S A 2006;103:5448–53.
28. Kinney TP, Merel N, Hart J, Joseph L, Waxman I. Microsatellite analysis of sporadic flat and depressed lesions of the colon. Dig Dis Sci 2005;50:327–30.