Published in final edited form as:
JCO Precis Oncol. 2017 ; 2017: .

Landscape of Microsatellite Instability Across 39 Cancer Types

Russell Bonneville, **Melanie A. Krook**, **Esko A. Kautto**, **Jharna Miya**, **Michele R. Wing**, **Hui-Zi Chen**, **Julie W. Reeser**, **Lianbo Yu**, and **Sameek Roychowdhury**
The Ohio State University, Columbus, OH.

Abstract

Purpose—Microsatellite instability (MSI) is a pattern of hypermutation that occurs at genomic microsatellites and is caused by defects in the mismatch repair system. Mismatch repair deficiency that leads to MSI has been well described in several types of human cancer, most frequently in colorectal, endometrial, and gastric adenocarcinomas. MSI is known to be both predictive and prognostic, especially in colorectal cancer; however, current clinical guidelines only recommend

Corresponding author: Sameek Roychowdhury, MD, PhD, The Ohio State University, 460 W 12th Ave, Room 508, Columbus, OH, 43210; sameek.roychowdhury@osumc.edu.
R.B. and M.A.K. contributed equally to this work.

The results published here, in whole or part, are based on data generated by The Cancer Genome Atlas managed by the National Cancer Institute (NCI) and National Human Genome Research Institute. Information about The Cancer Genome Atlas can be found at http://cancergenome.nih.gov.
The results published here, in whole or part, are based on data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative managed by the NCI. Information about TARGET can be found at http://ocg.cancer.gov/programs/target.

AUTHOR CONTRIBUTIONS
Conception and design: Russell Bonneville, Melanie A. Krook, Esko A. Kautto, Sameek Roychowdhury
Collection and assembly of data: Russell Bonneville, Melanie A. Krook, Sameek Roychowdhury
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST
The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or po.ascopubs.org/site/ifc.
Russell Bonneville
No relationship to disclose
Melanie A. Krook
No relationship to disclose
Esko A. Kautto
No relationship to disclose
Jharna Miya
No relationship to disclose
Michele R. Wing
No relationship to disclose
Hui-Zi Chen
No relationship to disclose
Julie W. Reeser
No relationship to disclose
Lianbo Yu
No relationship to disclose
Sameek Roychowdhury
Stock and Other Ownership Interests: Johnson & Johnson (I)
Research Funding: Takeda, Ignyta

MSI testing for colorectal and endometrial cancers. Therefore, less is known about the prevalence and extent of MSI among other types of cancer.

Methods—Using our recently published MSI-calling software, MANTIS, we analyzed whole-exome data from 11,139 tumor-normal pairs from The Cancer Genome Atlas and Therapeutically Applicable Research to Generate Effective Treatments projects and external data sources across 39 cancer types. Within a subset of these cancer types, we assessed mutation burden, mutational signatures, and somatic variants associated with MSI.

Results—We identified MSI in 3.8% of all cancers assessed—present in 27 of tumor types—most notably adrenocortical carcinoma (ACC), cervical cancer (CESC), and mesothelioma, in which MSI has not yet been well described. In addition, MSI-high ACC and CESC tumors were observed to have a higher average mutational burden than microsatellite-stable ACC and CESC tumors.

Conclusion—We provide evidence of as-yet-unappreciated MSI in several types of cancer. These findings support an expanded role for clinical MSI testing across multiple cancer types as patients with MSI-positive tumors are predicted to benefit from novel immunotherapies in clinical trials.

INTRODUCTION

Large-scale sequencing projects of cancer genomes have opened the door to studies that have identified putative biomarkers with potential clinical and therapeutic value, among them the presence or absence of microsatellite instability (MSI). Microsatellites are defined as 10 to 60 base pair regions that contain multiple repeats of 1 to 5 base pair motifs.[1] Microsatellites occur at microsatellite loci, which are widely dispersed throughout the human genome. In normal cells, repeat count of microsatellites is verified and maintained during cell division by the mismatch repair (MMR) system,[2,3] one of many cellular DNA repair mechanisms. Impairment of the MMR system can render cells unable to regulate the lengths of their microsatellites during cell division, termed MSI. After multiple cycles of cell division, cells with an impaired MMR system will develop varying lengths in their microsatellite sequences.

Mismatch repair deficiency is known to occur in some tumors,[2] either by somatic hypermutation of MMR genes, most commonly, *MLH1*[4,5]; an inherited germline MMR pathway mutation, such as in Lynch syndrome[6,7]; or double somatic mutations in MMR genes. MSI has been frequently observed within several types of cancer, most commonly in colorectal, endometrial, and gastric adenocarcinomas.[8,9] The clinical significance of MSI has been well described in colorectal cancer, as patients with MSI-H (MSI-high) colorectal tumors have been shown to have improved prognosis compared with those with MSS (microsatellite stable) tumors.[10,11] Furthermore, MSI-H colorectal tumors have been shown to be more susceptible to immune-enhancing therapies, such as the programmed cell death 1 (PD-1) inhibitor pembrolizumab,[12] which has been recently approved for any MSI-H or MMR-deficient unresectable or metastatic solid tumor.[13] Thus far, MSI-H tumors have the highest response rates to PD-1 inhibitors for any cancer type and have durable responses and a statistically significant improvement in overall survival.[12]

MSI polymerase chain reaction (PCR) and immunohistochemistry are two molecular biology– based methods that are in routine use for clinical MSI testing. MSI-PCR analyzes the distribution of microsatellite lengths at five standardized loci (Bethesda panel),[14] and immunohistochemistry detects the presence or absence of four proteins that are involved in the MMR pathway (*MSH2, MSH6, MLH1*, and *PMS2*). Recently, several computational methods have been developed that analyze next-generation sequencing (NGS) data to detect MSI. Examples of such software include mSINGS,[15] MSISensor,[16] and MANTIS.[17] A recent study by our group[17] demonstrated that MANTIS achieves high sensitivity (97%) and specificity (99%) across six cancer types—tested using samples with known MSI status by MSIPCR—and provides stable performance with varying numbers of microsatellite loci. Because of this, MANTIS is particularly well suited for application to a wider variety of cancer types.

As clinical MSI testing is routinely performed only on colorectal and endometrial tumors,[18] the prevalence of MSI in many other cancer types has been less well described. In addition, evidence exists that MSI-PCR may be less accurate in other cancer types.[19] A recent study by Hause et al[20] developed and applied the MSI detection tool, MOSAIC, to perform a detailed survey of MSI across 18 cancer types (n = 5,930 cases); however, many other cancer types have yet to be analyzed for MSI. The ability to detect MSI in novel cancer types would permit the investigation of immune-enhancing therapies in these cancers, with the potential to benefit previously unknown subsets of patients with cancer with MSI.

To perform a more comprehensive assessment of MSI across many additional cancer types than those analyzed by Hause et al, our study determined the prevalence of MSI in 39 distinct cancer types (n = 11,139 tumors from 11,080 patients) by using our previously published MSI-calling tool, MANTIS.

METHODS

Data Preprocessing—The Cancer Genome Atlas and Therapeutically Applicable Research to Generate Effective Treatments

For analysis, 10,701 cases of paired tumor-normal whole-exome sequencing data were obtained from The Cancer Genome Atlas (TCGA)[21–44] and Therapeutically Applicable Research to Generate Effective Treatments (TARGET)[45,46] projects. Data from all of these cases, with the exception of diffuse large B-cell lymphoma (DLBCL) were processed via our in-house automated pipeline, L-MAP (Landscape Microsatellite Analysis Pathway). L-MAP is implemented in Python and MySQL and was run on the Oakley supercomputer at the Ohio Supercomputing Center.[47] First, the metadata for all DNA whole-exome BAM files were downloaded from the Genomic Data Commons (GDC)[48] and were converted to SQL database entries. Aligned BAM files (to hg38[49]) were queried from GDC by LMAP by using the slicing end point provided by the GDC REST API. Reads that covered any base within 50 base pairs of a desired microsatellite locus were downloaded. As GDC data harmonization includes duplicate marking,[48] premarked duplicate reads were removed by using SAMtools (version 1.3.1).[50]

As a result of a GDC sample contamination issue, all 48 DLBCL paired tumor-normal cases were downloaded from the GDC Legacy Archive as whole-exome BAM files aligned to hg19 by using the GDC Data Transfer Tool. Premarked duplicate reads were removed as above.

Data Preprocessing—Other Sources

Four hundred thirty cases of paired tumor-normal whole-exome sequencing data were obtained from the Sequence Read Archive[51]: 338 chronic lymphocytic leukemia cases from 279 patients from Landau et al,[52] 32 cutaneous T-cell lymphoma cases from Choi et al,[53] 51 nasopharyngeal carcinoma cases from Zheng h et al,[54] and 8 cholangiocarcinoma cases from Ong et al.[55] Fifteen additional cholangiocarcinoma cases were obtained from the European Nucleotide Archive[56] from Chan-on et al.[57] All sample identifiers used are available in the Data Supplement. These cases were processed via L-MAP. Tumor and normal samples were downloaded in the FASTQ format using fastq-dump.[51] Alignment to hg38 was performed by using bwa (version 0.7.12)[58] with the mem algorithm. Duplicate reads were marked and removed by using Picard Mark- Duplicates.[59] Base quality score recalibration and indel realignment were performed by using GATK,[60] and the resulting BAM files were sliced, as above, by using SAM tools.

MSI Calling

MSI analysis with MANTIS (version 1.0.3; commit #942061f) was performed as previously described[17] for all cases by using an average distance threshold of 0.4 to differentiate MSI-H from MSS tumors. Coordinates for 2,539 microsatellite loci within or near the exome—originally introduced by Salipante et al[15] and used by later studies[17]— were converted from hg19 to hg38 by using Lift- Over.[61] Nine unlifted loci were discarded, which left 2,530 regions that were used for analysis with MANTIS in all cohorts, with the exception of DLBCL (Data Supplement). As the DLBCL data were aligned to hg19, the original 2,539 loci were used instead. MANTIS was run with author-recommended settings for whole-exome data—minimum read quality, 20; minimum locus quality, 25; minimum locus coverage, 20; minimum repeat reads, one; all other settings left at defaults. Eight samples were observed to have fewer than 10 loci sufficiently covered and were dropped. After MSI calling, microsatellite locus performance was assessed in each type of cancer as previously described. [17] Kernel density estimation functions were computed by using R (version 3.3.2) using the density() function with default settings.

Whole-Exome Analysis

For all tumor-normal pairs that were tested by MANTIS in adrenocortical carcinoma (ACC; n = 92), cervical cancer (CESC; n = 305), and mesothelioma (MESO; n = 83), we downloaded aligned reads from whole-exome sequencing. Reads were downloaded in BAM format from GDC by using the GDC Data Transfer Tool. Premarked duplicate reads were removed by using SAMtools,[50] variant calling was performed using MuTect[62] (see Variant Calling), and annotation was performed by using ANNOVAR (version 2016-02-01)[63] and GNU Parallel.[64]

Variant Calling

All variant calling was performed by using MuTect (version 1.1.7).[62] The target region was derived from RefSeq (release 80).[65] Exon data from the refGene table of the RefSeq Genes track was downloaded in BED format on February 28, 2017, by using the University of California, Santa Cruz Table Browser[66] and 100 base pair padding. Unknown contigs were excluded and overlapping regions were merged with BEDTools.[67] Variant cell format output was specified for MuTect and all other options were left at default. MuTect variant cell format output was then filtered for variants marked PASS. Variant annotation was performed by using ANNOVAR (version 2016- 02-01)[63] and GNU Parallel.[64] Somatic mutations in the repair genes *MSH2, MSH6, MLH1, PMS2, EXO1, POLD1*, and *POLE* were determined by filtering variants with a DANN[68,69] pathogenicity score greater than 0.96 (included in ANNOVAR). This threshold for DANN was chosen as it was previously shown to provide optimal sensitivity and specificity.[69]

Mutational signature calling was performed by using the tool deconstructSigs[70] with the Nature 2013 signatures set, which contains 27 signatures, [71] and the exome2genome normalization method. A mutational signature is a probability vector of length 96, with each element representing a single base change, along with bases immediately flanking it. In this analysis, linear regression is used to determine the relative contribution of each signature to the observed pattern of mutations. deconstructSigs was run over every ACC, CESC, and MESO sample by using all passing variants called with MuTect, as previously described.

All other downstream analyses were performed with Perl, Python, and R (version 3.3.2). Figures were generated by using R, Excel 2010 (Microsoft, Redmond, WA), and GraphPad Prism (version 7.0a; GraphPad Software, La Jolla, CA).

RESULTS

MSI Prevalence

We analyzed paired whole-exome sequencing data from 11,139 tumor-normal samples; 10,415 from the Cancer Genome Atlas (TCGA)[72] database, 280 from the TARGET[45] database, and 444 from other studies,[52–55,57] representing 39 distinct cancer types. MSI was detected in 27 of these 39 types of cancer (Fig 1A; Appendix Table A1; Data Supplement). The disease-specific prevalence of MSI varied widely, from 31.4% in endometrial carcinoma to 0.25% in glioblastoma multiforme. MSI was not detected in 12 cancer types (Figs 1A and 1B). Of 27 cancer types with MSI, 12 were found to have more than a single MSI-H tumor present and MSI-H prevalence greater than 1%. The relative level of instability, as measured by MANTIS score, varied substantially among MSI-H cancer types (Fig 1B and Appendix Fig A1 In addition, we attempted to determine which specific microsatellite loci performed best across the greatest number of cancer types (Data Supplement). Of 2,530 loci, we identified 22 loci that, within at least five cohorts, had an MSI-H versus MSS difference score greater than 0.75 and were sufficiently covered by at least 50% of samples in the cohort (Appendix Table A2). Only two loci that were assessed in the Bethesda[14] and Promega[73] MSI-PCR panels were included in our 2,530 loci, and neither of these were within the set of 22 top-performing loci. These results indicate a striking heterogeneity of

MSI patterns across various types of cancer. All four disease types with the highest rates of MSI prevalence were Lynchsyndrome–associated tumor types that have been previously known to exhibit MSI: endometrial carcinoma, colon adenocarcinoma, gastric adenocarcinoma, and rectal adenocarcinoma. Consistent with previous studies, MSI was observed to be more frequent in colon adenocarcinoma (19.7%) than rectal adenocarcinoma (5.7%).[20,74] Of importance, MSI was detected in three cancer types that have not been previously well characterized, most notably ACC (4.3%), cervical squamous cell carcinoma and CESC (2.6%), and MESO (2.4%; Fig 1A). To further investigate MSI status classifications, kernel density estimation[75,76] was performed on the MANTIS scores for these tumor types. This indicated clear distinctions between samples that MANTIS called MSI-H from samples called MSS (Fig 2). Kernel density estimation was also performed on all other tumor types tested (Appendix Fig A1).

Comparing Mutation Burden and Signatures Between MSI-H and MSS Tumors

As Lynch syndrome–associated MSI-H tumors have been shown to have higher somatic mutation burden,[12,77] we performed additional analyses to detect potential hypermutation in MSI-H ACC, CESC, and MESO. Somatic variant calling was performed on whole-exome samples from these four cancer types, and the mean absolute number of somatic mutations —both nonsynonymous and synonymous—was found to be increased among MSI-H versus MSS tumors within their own cohorts (Fig 3). In particular, an average of 1,157 somatic mutations were detected within MSI-H ACC samples versus 216 within MSS ACC (*P* = .01). An average of 5,675 somatic mutations were detected within MSI-H CESC samples versus 639 within MSS CESC (*P* = .003). Although statistical significance was not reached within MESO, MSI-H MESO tumors had, on average, a nearly seven-fold increase in mutational burden compared with MSS MESO tumors (982 *v* 142; *P* = .10). All *P* values were calculated by using Welch's two-sample *t* test with log normalization. These results indicate that MSI in ACC and CESC is correlated with high mutational burden.

To further investigate the observed somatic mutations in MSI-H versus MSS ACC, CESC, and MESO tumors, mutational signature analysis was performed by using a set of 27 signatures introduced by Alexandrov et al.[71] A mutational signature defines a pattern of preferential somatic mutation types and may be associated with a known biologic process or type of cancer. This analysis was first performed on pooled mutations among MSI-H or MSS samples within each of these three cancer cohorts (Appendix Fig A2). No clear pattern of signature differences was evident from this pooled analysis. Next, mutational signature analysis was performed for each individual case within these cohorts without pooling (Data Supplement). Differences among signature prevalence in ACC, CESC, and MESO did not reach statistical significance. *P* values were calculated by using two-sided Fisher's exact test (using signature presence or absence), with Benjamini correction for multiple hypotheses.[78]

MMR Pathway Alterations

MSI-H Lynch syndrome–associated tumors are known to lack the expression or function of at least one MMR protein; therefore, we analyzed somatic mutations that were predicted to be deleterious (by DANN[68]) in the MMR genes *MSH2, MSH6, MLH1, PMS2*, and *EXO1*, and the proofreading DNA polymerases *POLD1* and *POLE*, among MSI-H and MSS samples within ACC, CESC, and MESO (Appendix Table A3; Data Supplement). Although *POLD* and *POLE* are not considered MMR proteins, mutations in these genes have been shown to lead to somatic hypermutation.[22,79] Within these cohorts, 64% of MSI-H cases and 7% of MSS cases were found to contain at least one predicted deleterious somatic mutation in at least one of these genes; however, given that these samples were sequenced with potentially different exome captures, together with the increased mutational burden of MSI-H tumors, we could not determine the statistical significance of this finding.

DISCUSSION

In this study, we have performed, to our knowledge, the largest analysis of MSI in human cancer exomes to date, including 11,139 whole-exome tumor-normal pairs from 39 types of cancer. Compared with a study by Hause et al,[20] we observed similar rates of MSI in 18 types of cancer, and we also analyzed another 5,209 whole-exome tumor-normal pairs from 21 additional types of cancer. In addition, we observed that MSI-HACC and CESC tumors are significantly hypermutated compared with MSS ACC and CESC tumors. We identified three cohorts with significant MSI prevalence that have not been previously well described. Of particular interest, we identified MSI in 4 (4.4%) of 92ACCcases. Previous studies of MSI in ACC have implicated Lynch syndrome as a risk factor for familial ACC[80,81]; however, to our knowledge, NGS-based MSI analysis has not yet been applied to ACC.

MSI-H colorectal tumors have been previously shown to be exceptionally sensitive to therapy with PD-1 immune checkpoint inhibitors.[12] Identification of MSI in novel tumor types may lead to an expanded role for immunotherapy and a broader scope of clinical MSI testing.[82] In addition, MSI is known to be prognostic within colorectal cancer, [83] which may apply in other cancer types as well. For instance, Hause et al[20] provide evidence that increasing MSI positively correlates with survival time. Clinical trials of immune checkpoint inhibitors are beginning or are underway in ACC (ClinicalTrials.gov identifier: NCT02673333), CESC (ClinicalTrials.gov identifier: NCT02635360), and MESO (ClinicalTrials.gov identifiers: NCT02784171, NCT02991482, NCT02707666, and NCT02399371), and a previous study of dendritic cell immunotherapy in ACC[84] demonstrated tumor marker but not clinical response. These studies may benefit from the retrospective evaluation of MSI-H as a biomarker. Prospective expansion of clinical MSI testing to other cancer types may enlighten the prognostic and predictive value of MSI-H for noncolorectal cancers.

MMR deficiency is well recognized as the predominant cause of MSI within colorectal, endometrial, and gastric cancers. In addition, there have been anecdotal reports of ACC[80,81] as a potential extracolonic manifestation of Lynch syndrome. If future studies indicate that MSI in ACC, CESC, and/or MESO is indeed a result of MMR deficiency, the findings of this study may implicate previously unappreciated cancer types as being part of Lynch syndrome. Compared with germline alterations in MMR genes, somatic events are most often a result of hypermethylation of CpG islands in the promoter region of *MLH1*.[4] Additional investigation is needed to elucidate other molecular mechanisms that can lead to MSI, as well as the downstream effects of MSI on tumor-specific biology. In addition, of 9,569 tumors assessed in this study not within colorectal, endometrial, or gastric cancer, 77

(0.8%) were MSI-H. Only 14 of these were within ACC, CESC, or MESO, which compromised the statistical power of our mutational signature analysis. A larger cohort of MSI-H tumors would permit more comprehensive studies, including correlation with clinical data.

In summary, we have detected MSI in multiple cancer types, including ACC, CESC, and MESO, which indicates that MSI may affect non–Lynch syndrome tumor types. Within each type of cancer having MSI, we identified which loci—among 2,530—were most predictive of overall tumor MSI status. With additional analysis, these well-performing loci may form the basis of a targeted NGS panel for pancancer MSI detection. In addition, we found that MSI-H tumors in ACC and CESC have higher mutational burden than MSS tumors of these types. Given our observations of a long tail of MSI-H tumors across multiple cancer types, we propose that these and other, less common cancers undergo evaluation for MSI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank current and past members of the Roychowdhury laboratory for their helpful insight and discussion. Data used for this analysis are available at dbGaP (accession: phs000218. v17.p6). R.B. would like to dedicate this work to his late father, Russell E. Bonneville Jr.

Support

S.R. was supported by the American Cancer Society (Grant No. MRSG-12-194-01-TBG), the Prostate Cancer Foundation Young Investigator Award, the National Human Genome Research Institute (Grant No. UM1HG006508), the National Cancer Institute (Grant No. UH2CA202971), the American Lung Association, Pelotonia, and FORE Cancer Research; M.A.K. was supported by T32 Oncology Training Grant No. 5T32-CA009338; R.B. was supported by a university fellowship and M.R.W. was supported by the Helene Fuld Health Trust Nursing Scholarship. The chronic lymphocytic leukemia sequencing data (dbGaP: phs000922.v1.p1) used in this work was supported by National Human Genome Research Institute Large-Scale Sequencing Program Grant No. U54- HG003067 (to the Broad Institute).

References

1. Schlötterer C. Genome evolution: Are microsatellites really simple sequences? Curr Biol. 1998; 8:R132–R134. [PubMed: 9501977]
2. Shia J. Evolving approach and clinical significance of detecting DNA mismatch repair deficiency in colorectal carcinoma. Semin Diagn Pathol. 2015; 32:352–361. [PubMed: 25716099]
3. Strand M, Prolla TA, Liskay RM, et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 1993; 365:274–276. [Erratum: Nature 368:569, 1994]. [PubMed: 8371783]
4. Armaghany T, Wilson JD, Chu Q, et al. Genetic alterations in colorectal cancer. Gastrointest Cancer Res. 2012; 5:19–27. [PubMed: 22574233]
5. Kane MF, Loda M, Gaida GM, et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res. 1997; 57:808–811. [PubMed: 9041175]
6. Aaltonen LA, Peltomäki P, Leach FS, et al. Clues to the pathogenesis of familial colorectal cancer. Science. 1993; 260:812–816. [PubMed: 8484121]
7. Lynch HT, Shaw MW, Magnuson CW, et al. Hereditary factors in cancer. Study of two large midwestern kindreds. Arch Intern Med. 1966; 117:206–212. [PubMed: 5901552]

8. Imai K, Yamamoto H. Carcinogenesis and microsatellite instability: The interrelationship between genetics and epigenetics. Carcinogenesis. 2008; 29:673–680. [PubMed: 17942460]
9. Watson P, Lynch HT. The tumor spectrum in HNPCC. Anticancer Res. 1994; 14:1635–1639. [PubMed: 7979199]
10. Buckowitz A, Knaebel HP, Benner A, et al. Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases. Br J Cancer. 2005; 92:1746–1753. [PubMed: 15856045]
11. Benatti P, Gafà R, Barana D, et al. Microsatellite instability and colorectal cancer prognosis. Clin Cancer Res. 2005; 11:8332–8340. [PubMed: 16322293]
12. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015; 372:2509–2520. [PubMed: 26028255]
13. US Food and Drug Administration. Keytruda Biologics License Application 125514/S-14 approval letter. May 23, 2017. https://www.accessdata.fda.gov/drugsatfda_docs/appletter/2017/125514orig1s014ltr.pdf
14. Boland CR, Thibodeau SN, Hamilton SR, 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; 58:5248–5257. [PubMed: 9823339]
15. Salipante SJ, Scroggins SM, Hampel HL, et al. Microsatellite instability detection by next generation sequencing. Clin Chem. 2014; 60:1192–1199. [PubMed: 24987110]
16. Niu B, Ye K, Zhang Q, et al. MSIsensor: Microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014; 30:1015–1016. [PubMed: 24371154]
17. Kautto EA, Bonneville R, Miya J, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. Oncotarget. 2017; 8:7452–7463. [PubMed: 27980218]
18. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: A consensus statement by the US Multi-Society Task Force on colorectal cancer. Gastroenterology. 2014; 147:502–526. [PubMed: 25043945]
19. Faulkner RD, Seedhouse CH, Das-Gupta EP, et al. BAT-25 and BAT-26, two mononucleotide microsatellites, are not sensitive markers of microsatellite instability in acute myeloid leukaemia. Br J Haematol. 2004; 124:160–165. [PubMed: 14687025]
20. Hause RJ, Pritchard CC, Shendure J, et al. Classification and characterization of microsatellite instability across 18 cancer types. Nat Med. 2016; 22:1342–1350. [PubMed: 27694933]
21. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. [PubMed: 21720365]
22. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. [PubMed: 22810696]
23. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489:519–525. [Erratum: Nature 491:288, 2012]. [PubMed: 22960745]
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]
25. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368:2059–2074. [PubMed: 23634996]
26. Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. Nature. 2013; 497:67–73. [PubMed: 23636398]
27. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013; 499:43–49. [PubMed: 23792563]
28. Brennan CW, Verhaak RG, McKenna A, et al. The somatic genomic landscape of glioblastoma. Cell. 2013; 155:462–477. [Erratum: Cell 157:753, 2014]. [PubMed: 24120142]
29. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014; 507:315–322. [PubMed: 24476821]

30. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511:543–550. [Erratum: Nature 514:262, 2014]. [PubMed: 25079552]

31. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513:202–209. [PubMed: 25079317]

32. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014; 158:929–944. [PubMed: 25109877]

33. Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell. 2014; 26:319–330. [PubMed: 25155756]

34. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014; 159:676–690. [PubMed: 25417114]

35. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; 517:576–582. [PubMed: 25631445]

36. Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N Engl J Med. 2015; 372:2481–2498. [PubMed: 26061751]

37. Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. Cell. 2015; 161:1681–1696. [PubMed: 26091043]

38. Ciriello G, Gatza ML, Beck AH, et al. Comprehensive molecular portraits of invasive lobular breast cancer. Cell. 2015; 163:506–519. [PubMed: 26451490]

39. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of papillary renal-cell carcinoma. N Engl J Med. 2016; 374:135–145. [PubMed: 26536169]

40. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell. 2015; 163:1011–1025. [PubMed: 26544944]

41. Zheng S, Cherniack AD, Dewal N, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. Cancer Cell. 2016; 29:723–736. [Erratum: Cancer Cell 30:363, 2016]. [PubMed: 27165744]

42. Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017; 541:169–175. [PubMed: 28052061]

43. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017; 543:378–384. [PubMed: 28112728]

44. Fishbein L, Leshchiner I, Walter V, et al. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. Cancer Cell. 2017; 31:181–193. [PubMed: 28162975]

45. National Cancer Institute. TARGET: Therapeutically Applicable Research to Generate Effective Treatments. https://ocg.cancer.gov/programs/target

46. Pugh TJ, Morozova O, Attiyeh EF, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013; 45:279–284. [PubMed: 23334666]

47. Ohio Supercomputer Center. Oakley. https://www.osc.edu/resources/technical_support/supercomputers/oakley

48. Grossman RL, Heath AP, Ferretti V, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016; 375:1109–1112. [PubMed: 27653561]

49. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [Erratum: Nature 411:720, 2001]. [PubMed: 11237011]

50. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

51. Leinonen R, Sugawara H, Shumway M. The sequence read archive. Nucleic Acids Res. 2011; 39:D19–D21. [PubMed: 21062823]

52. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. Nature. 2015; 526:525–530. [PubMed: 26466571]

53. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. Nat Genet. 2015; 47:1011–1019. [PubMed: 26192916]

54. Zheng H, Dai W, Cheung AKL, et al. Whole-exome sequencing identifies multiple loss-of-function mutations of NF-kB pathway regulators in nasopharyngeal carcinoma. Proc Natl Acad Sci USA. 2016; 113:11283–11288. [PubMed: 27647909]

55. Ong CK, Subimerb C, Pairojkul C, et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. Nat Genet. 2012; 44:690–693. [PubMed: 22561520]

56. Leinonen R, Akhtar R, Birney E, et al. The European Nucleotide Archive. Nucleic Acids Res. 2011; 39:D28–D31. [PubMed: 20972220]

57. Chan-On W, Nairismägi M-L, Ong CK, et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. Nat Genet. 2013; 45:1474–1478. [PubMed: 24185513]

58. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

59. Broad Institute. Picard tools. http://broadinstitute.github.io/picard

60. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]

61. Hinrichs AS, Karolchik D, Baertsch R, et al. The UCSC Genome Browser Database: Update 2006. Nucleic Acids Res. 2006; 34:D590–D598. [PubMed: 16381938]

62. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–219. [PubMed: 23396013]

63. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]

64. Usenix. GNU Parallel—The command-line power tool. https://www.usenix.org/system/files/login/articles/105438-Tange.pdf

65. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44:D733–D745. [PubMed: 26553804]

66. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32:D493–D496. [PubMed: 14681465]

67. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]

68. Quang D, Chen Y, Xie X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 2015; 31:761–763. [PubMed: 25338716]

69. Jensen, D. The best variant prediction method that no one is using. http://www.enlis.com/blog/2015/03/17/the-best-variant-prediction-method-that-no-one-is-using/

70. Rosenthal R, McGranahan N, Herrero J, et al. DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016; 17:31. [PubMed: 26899170]

71. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421. [Erratum: Nature 502:502, 2013]. [PubMed: 23945592]

72. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas pan-cancer analysis project. Nat Genet. 2013; 45:1113–1120. [PubMed: 24071849]

73. Bacher JW, Flanagan LA, Smalley RL, et al. Development of a fluorescent multiplex assay for detection of MSI-High tumors. Dis Markers. 2004; 20:237–250. [PubMed: 15528789]

74. Phipps AI, Lindor NM, Jenkins MA, et al. Colon and rectal cancer survival by tumor location and microsatellite instability: The Colon Cancer Family Registry. Dis Colon Rectum. 2013; 56:937–944. [PubMed: 23838861]

75. Parzen E. On estimation of a probability density function and mode. Ann Math Stat. 1962; 33:1065–1076.

76. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Stat. 1956; 27:832–837.

77. Gatalica Z, Vranic S, Xiu J, et al. High microsatellite instability (MSI-H) colorectal carcinoma: A brief review of predictive biomarkers in the era of personalized medicine. Fam Cancer. 2016; 15:405–412. [PubMed: 26875156]

78. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B (Methodological). 1995; 57:289–300.

79. Johnson RE, Klassen R, Prakash L, et al. A major role of DNA polymerase δ in replication of both the leading and lagging DNA strands. Mol Cell. 2015; 59:163–175. [PubMed: 26145172]

80. Challis BG, Kandasamy N, Powlson AS, et al. Familial adrenocortical carcinoma in association with Lynch syndrome. J Clin Endocrinol Metab. 2016; 101:2269–2272. [PubMed: 27144940]

81. Raymond VM, Everett JN, Furtado LV, et al. Adrenocortical carcinoma is a Lynch syndrome-associated cancer. J Clin Oncol. 2013; 31:3012–3018. [PubMed: 23752102]

82. Dudley JC, Lin MT, Le DT, et al. Microsatellite instability as a biomarker for PD-1 blockade. Clin Cancer Res. 2016; 22:813–820. [PubMed: 26880610]

83. Kawakami H, Zaanan A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. Curr Treat Options Oncol. 2015; 16:30. [PubMed: 26031544]

84. Papewalis C, Fassnacht M, Willenberg HS, et al. Dendritic cells as potential adjuvant for immunotherapy in adrenocortical carcinoma. Clin Endocrinol (Oxf). 2006; 65:215–222. [PubMed: 16886963]

APPENDIX

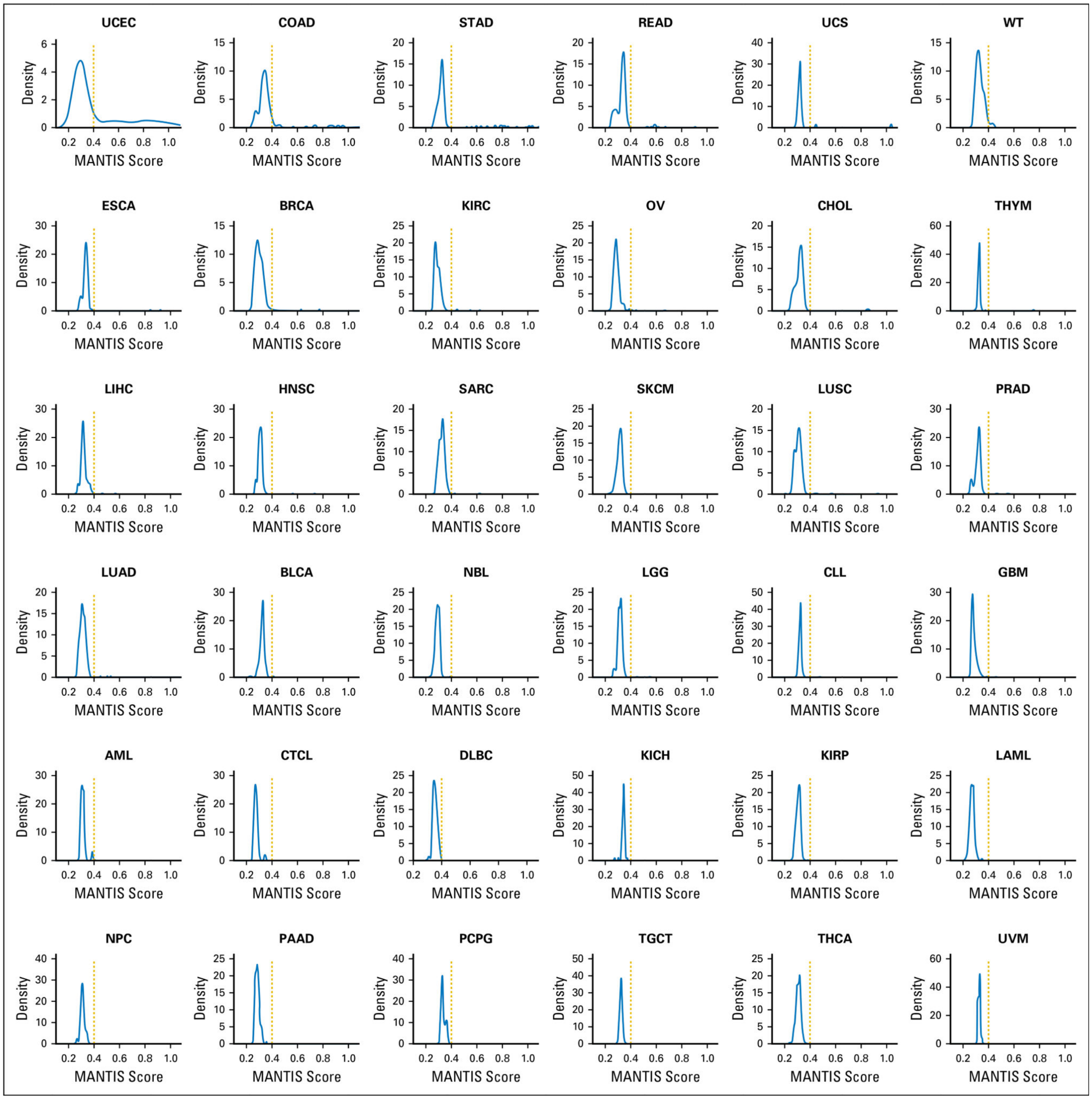

Fig. A1.

Kernel density plots of MANTIS scores within 36 cancer types. The dotted line denotes the average distance threshold of 0.4, used by MANTIS to differentiate microsatellite instability high from microsatellite stable tumors. Uterine corpus endometrial carcinoma (UCEC): kernel bandwidth (h) = 4.89e-02. Colon adenocarcinoma (COAD): h = 1.13e-02. Stomach adenocarcinoma (STAD): h = 7.59e-03. Rectal adenocarcinoma (READ): h = 9.16e-03. Uterine carcinosarcoma (UCS): h = 4.10e-03. Pediatric high-risk Wilms tumor (WT): h = 1.27e-02. Esophageal carcinoma (ESCA): h = 5.02e-03. Breast carcinoma (BRCA): h =

7.41e-03. Kidney renal clear cell carcinoma (KIRC): h = 6.83e-03. Ovarian serous cystadenocarcinoma (OV): h = 5.23e-03. Cholangiocarcinoma (CHOL): h = 1.17e-02. Thymoma (THYM): h = 3.08e-03. Liver hepatocellular carcinoma (LIHC): h = 4.42e-03. Head and neck squamous cell carcinoma (HNSC): h = 4.25e-03. Sarcoma (SARC): h = 7.14e-03. Skin cutaneous melanoma (SKCM): h = 5.32e-03. Lung squamous cell carcinoma (LUSC): h = 7.13e-03. Prostate adenocarcinoma (PRAD): h = 5.31e-03. Lungadenocarcinoma (LUAD:): h = 5.74e-03. Bladder carcinoma (BLCA): h = 4.40e-03. Pediatric neuroblastoma (NBL:): h = 5.47e-03. Lower-grade glioma (LGG:): h = 4.32e-03. Chronic lymphocytic leukemia (CLL): h = 2.64e-03. Glioblastoma multiforme (GBM): h = 4.38e-03.Pediatric acute myeloid leukemia (AML): h = 6.13e-03. Cutaneous T-cell lymphoma(CTCL): h = 5.86e-03. Diffuse large B-cell lymphoma(DLBC): h = 6.68e-03. Kidney chromophobe (KICH): h = 3.34e-03. Kidney renal papillary cell carcinoma (KIRP): h = 5.16e-03. Acute myeloid leukemia (LAML): h = 5.28e-03. Nasopharyngeal carcinoma (NPC): h = 6.09e-03. Pancreatic adenocarcinoma (PAAD): h = 5.36e-03. Pheochromocytoma and paraganglioma (PCPG): h = 5.04e-03. Testicular germ cell tumor (TGCT): h = 3.40e-03. Thyroid carcinoma (THCA): h = 5.09e-03. Uveal melanoma (UVM): h = 3.06e-03.

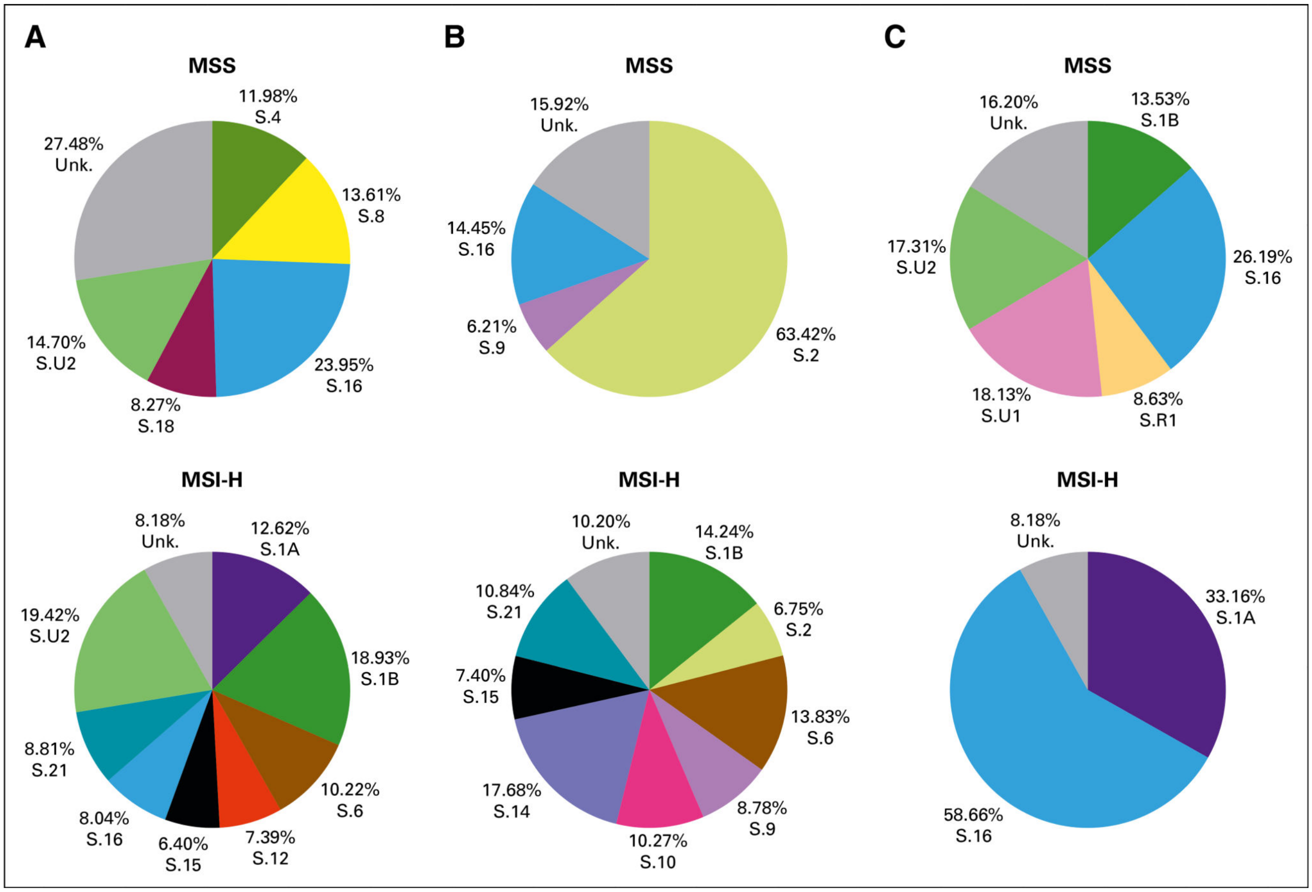

Fig. A2.

Patterns of mutational signatures (S) across microsatellite instability cancers: (A) adrenocortical carcinoma (ACC), (B) cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), and (C) mesothelioma (MESO). Mutational signatures were called using deconstructSigs from pooled variants from all microsatellite instability high or microsatellite stable tumors within each cohort within ACC, CESC, and MESO. Unk., unknown.

Table A1

Summary of MSI Landscape Analysis

Cancer Type	No. of Cases	MSI-H	% MSI-H
Adrenocortical carcinoma (TCGA-ACC)	92	4	4.35
Bladder carcinoma (TCGA-BLCA)	412	2	0.49
Breast carcinoma (TCGA-BRCA)	1,044	16	1.53
Cervical squamous cell carcinoma and endocervical adenocarcinoma (TCGA-CESC)	305	8	2.62
Cholangiocarcinoma (TCGA-CHOL, CHOL_10.1038_ng.2273, CHOL_10.1038_ng.2806)	74	1	1.35
Chronic lymphocytic leukemia (CLL_phs000922.v1.p1)	338	1	0.30
Colon adenocarcinoma (TCGA-COAD)	431	85	19.72
Cutaneous T-cell lymphoma (CTCL_10.1038_ng.3356)	33	0	0.00
Lymphoid neoplasm diffuse large B-cell lymphoma (TCGA-DLBC)	48	0	0.00
Esophageal carcinoma (TCGA-ESCA)	184	3	1.63
Glioblastoma multiforme (TCGA-GBM)	396	1	0.25
Head and neck squamous cell carcinoma (TCGA-HNSC)	510	4	0.78
Kidney chromophobe (TCGA-KICH)	66	0	0.00
Kidney renal clear cell carcinoma (TCGA-KIRC)	339	5	1.47
Kidney renal papillary cell carcinoma (TCGA-KIRP)	288	0	0.00
Acute myeloid leukemia (TCGA-LAML)	146	0	0.00
Lower-grade glioma (TCGA-LGG)	513	2	0.39
Liver hepatocellular carcinoma (TCGA-LIHC)	375	3	0.80
Lung adenocarcinoma (TCGA-LUAD)	569	3	0.53
Lung squamous cell carcinoma (TCGA-LUSC)	496	3	0.60
Mesothelioma (TCGA-MESO)	83	2	2.41
Nasopharyngeal carcinoma (NPC_10.1073_pnas.1607606113)	50	0	0.00
Ovarian serous cystadenocarcinoma (TCGA-OV)	437	6	1.37
Pancreatic adenocarcinoma (TCGA-PAAD)	183	0	0.00
Pheochromocytoma and paraganglioma (TCGA-PCPG)	179	0	0.00
Prostate adenocarcinoma (TCGA-PRAD)	498	3	0.60
Rectal adenocarcinoma (TCGA-READ)	157	9	5.73
Sarcoma (TCGA-SARC)	255	2	0.78
Skin cutaneous melanoma (TCGA-SKCM)	470	3	0.64
Stomach adenocarcinoma (TCGA-STAD)	440	84	19.09
Testicular germ cell tumor (TCGA-TGCT)	150	0	0.00
Thyroid carcinoma (TCGA-THCA)	496	0	0.00

Cancer Type	No. of Cases	MSI-H	% MSI-H
Thymoma (TCGA-THYM)	123	1	0.81
Uterine corpus endometrial carcinoma (TCGA-UCEC)	542	170	31.37
Uterine carcinosarcoma (TCGA-UCS)	57	2	3.51
Uveal melanoma (TCGA-UVM)	80	0	0.00
Pediatric acute myeloid leukemia (TARGET-AML)	19	0	0.00
Pediatric neuroblastoma (TARGET-NBL)	220	1	0.45
Pediatric high-risk Wilms tumor (TARGET-WT)	41	1	2.44
Total	11,139	425	3.82

NOTE. Listed for each cancer type are the number of cases analyzed and those called MSI-H by MANTIS. Note that for CLL, these 338 cases were from 279 patients, many of whom had multiple tumor samples.

Abbreviations: MSI, microsatellite instability; MSI-H, microsatellite instability high; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; TCGA, The Cancer Genome Atlas.

Table A2

All Microsatellite Loci With Difference Scores of > 0.75 in Five or More Cancer Types

Locus	Count	Cancer Type	K-mer
chr5: 14485053-14485065	8	BRCA, CHOL, COAD, ESCA, LUSC, STAD, THYM, UCEC	(T)13
chr13: 27559820-27559834	7	COAD, ESCA, GBM, READ, STAD, UCEC, UCS	(A)15
chr13: 78642222-78642234	7	COAD, ESCA, LGG, STAD, THYM, UCEC, UCS	(A)13
chr8: 102275623-102275635	7	CHOL, COAD, ESCA, LUSC, STAD, THYM, UCEC	(A)13
chr18: 62275354-62275366	6	CHOL, COAD, GBM, LGG, LUSC, STAD	(T)13
chr3: 140959543-140959557	6	ACC, CHOL, COAD, ESCA, READ, UCEC	(A)15
chr6: 152419547-152419559	6	ACC, CHOL, COAD, ESCA, OV, READ	(A)13
chr7: 93271201-93271214	6	NBL, CHOL, COAD, READ, STAD, UCS	(T)14
chr1: 230958305-230958320	5	CHOL, COAD, ESCA, STAD, THYM	(A)16
chr1: 31915992-31916005	5	WT, CHOL, ESCA, SARC, THYM	(A)14
chr1: 77966823-77966836	5	COAD, LUSC, READ, STAD, UCEC	(A)14
chr14: 30722463-30722475	5	CHOL, ESCA, LIHC, LUSC, STAD	(T)13
chr2: 119956826-119956841	5	CHOL, COAD, ESCA, GBM, STAD	(T)16
chr2: 200913995-200914009	5	CHOL, COAD, ESCA, GBM, UCEC	(A)15
chr20: 38517489-38517502	5	NBL, CHOL, ESCA, STAD, UCEC	(T)14
chr3: 112155056-112155069	5	CHOL, COAD, ESCA, PRAD, STAD	(A)14
chr4: 38132803-38132818	5	CHOL, COAD, ESCA, READ, THYM	(T)16
chr5: 53062932-53062944	5	CHOL, OV, PRAD, THYM, UCS	(A)13
chr6: 111008019-111008035	5	CHOL, COAD, ESCA, STAD, UCEC	(T)17
chr7: 74753041-74753054	5	COAD, ESCA, STAD, UCEC, UCS	(A)14
chr8: 129862369-129862381	5	ESCA, READ, STAD, THYM, UCS	(A)13
chr9: 99968416-99968429	5	ESCA, GBM, LUSC, SARC, THYM	(T)14

NOTE. A locus was only considered in a cancer type if sufficient sequencing coverage of the locus was present in at least 50% of cases in that cancer type, including at least one microsatellite instability high sample.

Abbreviations: ACC, adrenocortical carcinoma; BRCA, breast carcinoma; CHOL, cholangiocarcinoma; chr, chromosome; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; LGG, lower-grade glioma; LIHC, liver hepatocellular carcinoma; LUSC, lung squamous cell carcinoma; NBL, neuroblastoma; OV, ovarian serous cystadenocarcinoma; PRAD, prostate adenocarcinoma; READ, rectal adenocarcinoma; SARC, sarcoma; STAD, stomach adenocarcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; WT, Wilms tumor.

Table A3

Frequency of Predicted Deleterious MMR Mutations in ACC, CESC, and MESO

Variable	Total No. of Samples	*MSH2*	*MSH6*	*MLH1*	*PMS2*	*EXO1*	*POLE*	Total No. of Samples With at Least One Predicted Deleterious Mutation
ACC								
MSS	88	1	1	0	1	0	1	4
MSI-H	4	0	0	1	0	0	1	2
CESC								
MSS	297	3	3	5	0	3	10	22
MSI-H	8	0	1	3	1	1	2	6
MESO								
MSS	81	0	1	0	0	0	1	2
MSI-H	2	1	0	0	0	0	0	1
ACC + CESC + MESO								
MSS	466	4	5	5	1	3	12	28
MSI-H	14	1	1	4	1	1	3	9

NOTE. Listed are the number of samples (MSS or MSI-H) with at least one predicted deleterious mutation in *MSH2, MSH6, MLH1, PMS2, EXO1, POLD1*, and *POLE*. Mutations were called by using MuTect ("Variant Calling" in Methods) and included in this table if the DANN pathogenicity score was > 0.96.

Abbreviations: ACC, adrenocortical carcinoma; CESC, cervical cancer; MESO, mesothelioma; MMR, mismatch repair; MSI, microsatellite instability; MSI-H, microsatellite instability high; MSS, microsatellite stable.

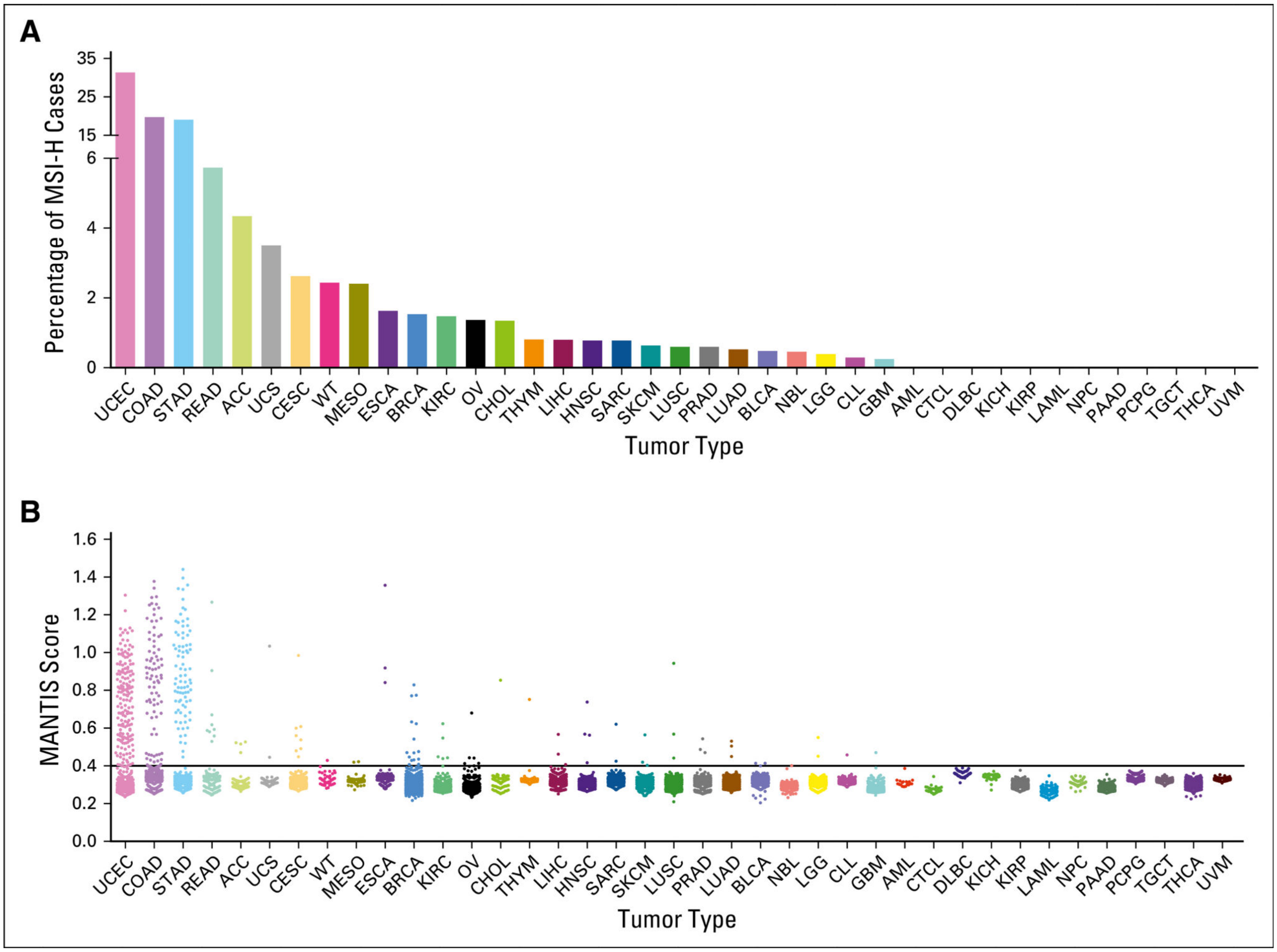

Fig 1.

Prevalence of microsatellite instability (MSI) across 39 human cancer types. (A) MSI prevalence was detected across 39 tumor types. The total number of tumors and the percentage of cases called MSI-high (MSI-H) in each cohort is listed in Appendix Table A1. (B) The relative level of instability, as measured by MANTIS score, is shown across all 39 tumor types. Note that for chronic lymphocytic leukemia (CLL), the listed MSI prevalence in panel A is out of 279 patients, and all 338 tumors are shown in panel B. MANTIS threshold cutoff of 0.4 is depicted with a dashed line. ACC, adrenocortical carcinoma; AML, pediatric acute myeloid leukemia (TARGET); BLCA, bladder carcinoma; BRCA, breast carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; CTCL, cutaneous T-cell lymphoma; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia (TCGA); LGG, lower-grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; NBL, pediatric neuroblastoma; NPC, nasopharyngeal carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma;

PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectal adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TCGT, testicular germ cell tumor; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma; WT, Wilms tumor.

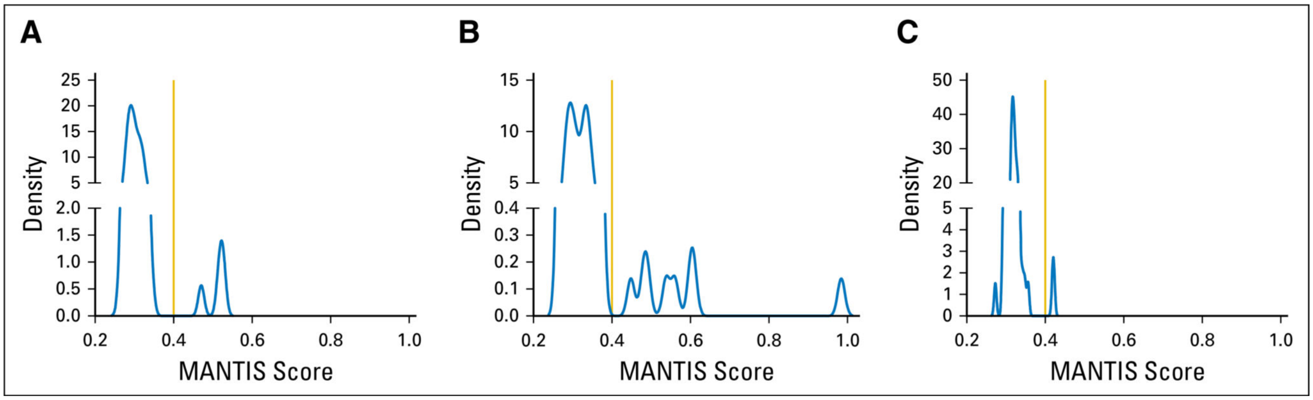

Fig 2.

Kernel density plots of MANTIS scores within (A) adrenocortical carcinoma (ACC), (B) cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), and (C) mesothelioma (MESO). The dotted line denotes the average distance threshold of 0.4, used by MANTIS to differentiate microsatellite instability high from microsatellite stable tumors. ACC: n = 92, kernel bandwidth (h) = 7.6e-3; CESC: n = 305, h = 9.4e-3; MESO: n = 83, h = 3.2e-3. KD plots for the other 36 cancer types analyzed are available in Appendix Fig A1.

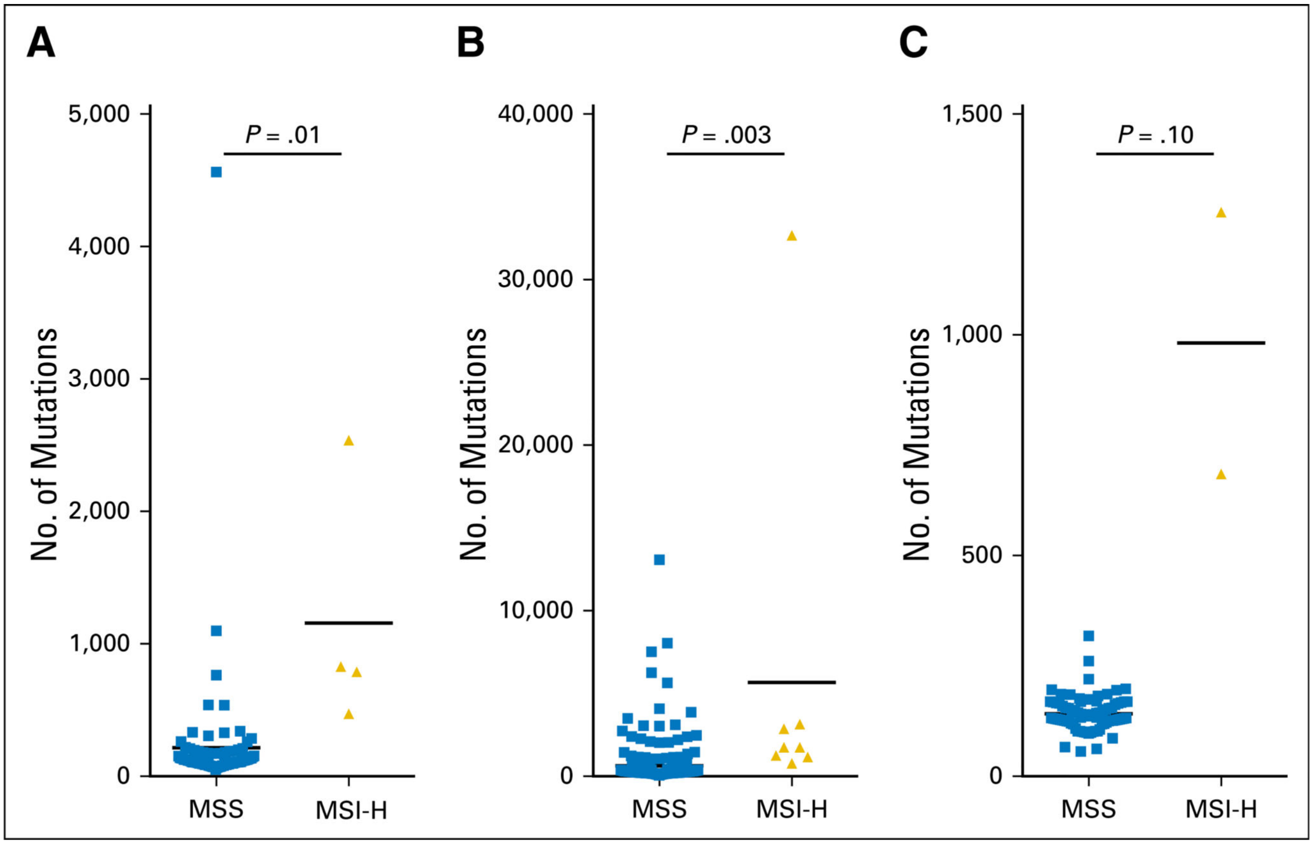

Fig 3.

Somatic mutational burden correlates with microsatellite instability high (MSI-H) status within adrenocortical carcinoma (ACC) and cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC). Mutational burden is listed for (A) ACC, (B) CESC, and (C) mesothelioma (MESO). *P* values were calculated using the Welch two-sample *t* test of log-normalized absolute somatic mutation counts. Variant calling was performed by using MuTect (“Variant Calling” in Methods), and all passing variants were included (nonsynonymous or synonymous).