Supplementary Material

Supplementary Methods

Supplementray Tables

Supplementary Table 1. PCR primers used for fragment size analysis (n = 30)

Supplementary Table 2. Thermal cycling conditions per PCR fragment size analysis and Sanger sequencing primer pair

Supplementary Table 3. PCR conditions for GSP1 and GSP2 amplifications of all panels using the ArcherDX LiquidPlex Library Prep Kit

Supplementary Table 4. Targets utilized in Archer NGS sequencing (n = 168)

Supplementary Table 5. Summary of FSM detection in supernatant DNA of MSI-H cells via PCR fragment size analysis

Supplementary Table 6. Number of FSMs detected and detection rate in a DNA spike-in assay

Supplementary Table 7. The lowest spike-in samples in which FSMs were detected*

Supplementary Table 8. Summary of FSM detection in cfDNA from healthy participants

Supplementary Table 9. Number of healthy cfDNA with FSMs detected

Supplementary Table 10. Depth and DAF of several targets in matched tumor and cfDNA samples

Supplementary Table 11. Number of FSMs detected in matched MSI-H tumor, cfDNA, buffy coat, and adjacent normal tissue with patient clinical information

Supplementray Figures

Supplementary Figure 1. Diagram of patient samples used in the study.

Supplementary Figure 2. Flow diagram to show the sources of DNA (including DNA tissue sources used to generate a normal dataset for background error correction for NGS data analysis) and platforms used at the various panel development stages.

Supplementary Figure 3. DNA size analysis by Agilent TapeStation system. Lane 1: DNA size marker; Lane 2 and 3: DNA from LoVo and HCT116 culture supernatants; lane 4 and 5: cellular DNA from HCT116 and LoVo; and lane 6, 7, and 8: cfDNA from plasma_EDTA, plasma_ACD and serum from healthy participants.

Supplementary Figure 4. Number of FSMs detected in **(A)** cfDNA and **(B)** tumors by MMR gene. MMR status was performed by immunohistochemical staining and data was not avialable for some samples. There was no significant difference among groups for either cfDNA or tumors. However, several limitations should be considered when interpreting the data (e.g., small small size, low and different plasma or serum volume used from each sample, different tumor stage and grade, etc.).

Supplementary Methods

Cell culture

Microsatellite instability high (MSI-H) colorectal cancer (CRC) cell lines HCT116, LoVo, KM12, and HCT15 and microsatellite stable (MSS) CRC cell lines Colo205 and HT29 were obtained from the Division of Cancer Treatment and Diagnosis (DCTD) Tumor Repository at National Cancer Institute (NCI)-Frederick (https://dtp.cancer.gov/repositories/DCTDTumorRepository/, Frederick, MD). They were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS), 2mM L-Glutamine, and 1% penicillin-streptomycin (P/S). MSI-H CRC cell line RKO was obtained from the American Type Culture Collection (ATCC; Manassas, Virginia) and cultured in ATCC-formulated Eagle's Minimum Essential Medium with 10% FBS. All cells were incubated at 5% CO_2 at 37ºC. All cell lines tested mycoplasma negative. Cell lines were authenticated either by the ATCC or by the DCTD using Applied Biosystems AmpFLSTR™ Identifiler™ PCR amplification kit (ThermoFisher Scientific, Waltham, MA) prior to cell line receipt.

Supernatant or cellular DNA extraction from cultured cells

DNA extraction from cell culture supernatant was described by Fricke et al. (1). Briefly, CRC cell lines were grown in complete media as described above in T150 flasks until they reached 80-90% confluency, then washed with PBS twice, and cultured for 16 hours in 17 mL of serum- and antibiotic-free media. Media were collected on the second day and subjected to sequential low speed centrifugations (480 g and 2000 g at 4°C for 10min). The supernatants were then passed through a 0.22uM filter and concentrated using 10KDa-cutoff Vivaspin 20 concentrators (4000 g at 4°C for 30min; Millipore Sigma, MO). The concentrates were further processed using Invitrogen Total Exosome Isolation kit (from cell culture media) (ThermoFisher Scientific, Grand Island, NY). Cells were also collected at the same time for cellular DNA isolation. Supernatant and cellular DNA was extracted using QIAamp DNA mini kit (Qiagen, Germantown, MD). The mean supernatant DNA concentration was 13.7 ng/uL for HCT116, 21.6 ng/uL for LoVo, 61.3 ng/uL for HT29, 8.2 ng/uL for KM12, and 3.9 ng/uL for RKO from several batches of supernatants.

Blood sample processing

Blood acquired through the National Institute of Health (NIH) Clinical Center (Bethesda, Maryland) was collected into EDTA vacutainer tubes and transported to Fredrick National Lab for Cancer Research (FNLCR) on wet ice. Upon arrival, blood samples were spun for 10 minutes at 500 g. Plasma on the top layer was transferred to a 15 mL conical tube and spun at 2000 g for 10 minutes. Plasma was stored in 0.5 mL aliquots at -80ºC until DNA extraction. Blood from other sources were processed by each site and plasma was stored at -80ºC before shipping to FNLCR on dry ice.

DNA extraction from patient samples

DNA from tumors, buffy coat, or adjacent normal tissue was extracted by each collection sites. cfDNA was extracted from 2-8 mL of plasma from healthy participants or 0.5-2 mL of

plasma from MSI-H and MMRd patients using Quick-cfDNA/cfRNA Serum & Plasma Kit (Zymo Research, Irvine, CA) or Qiagen circulating nucleic acid kit (Qiagen, Germantown, MD). DNA or cfDNA was quantified using Qubit Fluorometer (ThermoFisher Scientific, Waltham, MA). DNA fragment size was analyzed using DNA ScreenTape, Cell-free DNA ScreenTape, or Genomic DNA ScreenTape, as appropriate, on Agilent 2200 TapeStation system (Agilent, Santa Clara, CA).

PCR fragment size analysis

Two small panels (12-target and 30-target) were initially tested using PCR-based fragment size analysis using 5 ng of supernatant DNA for each target as described previously (2) on ABI 3730xl DNA Analyzer (ThermoFisher Scientific, Waltham, MA). PCR amplification of cMNRs sequences was performed using FAM-labeled PCR primers (**Supplementary Tables 1 and 2**). PCR products were run on Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) before injecting into DNA Analyzer. Data was analyzed using Genescan Analysis software for fragment size analysis and validated by Sanger sequencing using ABI 3730xl DNA Analyzer (ThermoFisher Scientific) as described previously (2).

Panel development and targeted sequencing via Archer Next Generation Sequencing (NGS) platform

To enrich the target templates and increase the detection sensitivity, primers were custom designed using Anchored Multiplex PCR (AMP™) chemistry developed by Archer (now Integrated DNA Technologies (IDT), Newark, NJ). All samples were processed using the ArcherDx LiquidPlex Library Prep Kit and LiquidPlex ctDNA Protocol for Illumina (Archer) and PCR conditions listed in **Supplementary Table 3**. Prior to sequencing, all libraries were quantitated on QuantStudio DX (ThermoFisher Scientific) using the KAPA Universal Library Quantification Kit for Illumina (Roche, Basel, Switzerland) and standard protocol. The initial 12-target panel was tested on MiSeq DX using a MiSeq Reagent Kit v2 (300 cycles; Illumina, San Diego, CA) and 48 ng spike-in samples of mutant DNA from HCT116 supernatant into WT DNA from HT29 supernatant at 5%, 1%, 0.5%, 0.25%, 0.1%, and 0.05% to test detection capabilities. Seven libraries with 100% WT DNA from HT29 supernatant were also sequenced and used as a control dataset for background error correction for data analysis. Optimization of data output was carried out to increase the sensitivity by migrating sequencing to NextSeq 550 (Illumina) using the NextSeq 500/550 Mid Output Kit v2.5 (300 cycles; Illumina), increasing library prep input concentration to 100 ng, decreasing PhiX loading percentage from 20% to 10% due to preferential flow cell binding, and adding a preliminary fragmentation step to library prep to increase unique target sites. A panel with 30-target was initially tested, then expanded to include additional 138 targets (total 168 targets). Spike-in studies were repeated at optimal data output conditions with 100 ng input concentrations, as well as expanded to include spike-in samples utilizing DNA from KM12 and LoVo cell lines, each respectively created at the spike-in percentages of 20%, 10%, 5%, 1%, 0.5%, 0.25%, 0.1%, and 0.05% using WT DNA from HT29. Following panel expansion, sequencing output was increased further by switching to the NextSeq 500/550 High Output Kit v2.5 (300 cycles; Illumina). Targets were culled to establish a final panel consisting of only the 122 targets found to show detection at ≤ 1% spike-in samples (**Supplementary Table 4**).

FSM detection in MSI-H MMRd patients and healthy participants using 122-target panel and Archer NextSeq platform

Following 122-target panel optimization, all the matched tumor, adjacent normal, and buffy coat samples were processed as stated above using 50 ng starting input with fragmentation preceding library preparation. For plasma of MSI-H MMRd carrier or patients, all the cfDNA extracted from 0.5-2 mL plasma (1-50 ng for each sample) was used for library construction using Archer reagents. cfDNA yield was very low in most of healthy participants (< 5 ng). Thus, some of them were pooled and 20-50 ng was used as input for library construction. In addition, seven buffy coat DNA or pooled cfDNA samples from healthy participants were also sequenced for background error correction. The 122-target panel was utilized, and amplification conditions are stated in **Supplementary Table 3**. Sequencing was performed on NextSeq 550 using the NextSeq 500/550 High Output Kit v2.5 (300 cycles).

NGS data analysis

Raw FASTQ files from the Illumina sequencer were uploaded to the Archer® Analysis Virtual Machine with a NGS data analysis pipeline developed by Archer (Archer Analysis Suite v6.2.7, Archer, now IDT) for data analysis. For background error correction, three different normal datasets generated from supernatant DNA from MSS cell line HT29, and buffy coat or cfDNA from healthy participants were used for the analysis of data generated from different DNA sample sources (supernatant DNA from MSI-H cell lines, and tumor, adjacent normal tissue, buffy coat, or cfDNA from LS patients). The number of alternate observations (AO) for a variant call from deep (i.e., error-correctable) molecular bins (DAO) and the total sequence coverage at this position from deep (i.e., error-correctable) molecular bins (DDP) were used to calculate DAF (the allele fraction of the reads from deep (i.e., error-correctable) molecular bins that support the alternative allele; DAO/DDP). ND DAF (Normal Dataset Detectable Allele Fraction) Outlier P-value (the probability this mutation was due to background noise given the provided Normal Dataset and unique molecules with a deep amplicon depth into account) was computed and positive mutation calling was made when $P < .05$.

Receiver-Operating Characteristic (ROC) analysis

ROC analysis was performed to evaluate the performance of the panel in distinguishing MSI-H and MMRd patients from healthy participants using the number of FSMs detected in each patient or healthy participant. MSI-H and MMRd patients were defined as positive and healthy participants as negative. To determine whether the performance can be improved, ROC analysis was modeled by removing those targets detected in healthy participants and further removing those detected in low number of patients. Data was graphed using GraphPad Prism 10.0.2 (GraphPad Software, LLC).

Reference

1. Fricke F, Lee J, Michalak M*, et al.* TGFBR2-dependent alterations of exosomal cargo and functions in DNA mismatch repair-deficient HCT116 colorectal cancer cells. *Cell Commun Signal.* 2017;15(1):14. doi:10.1186/s12964-017-0169-y.
2. Song Y, Kerr TD, Sanders C*, et al.* Organoids and metastatic orthotopic mouse model for mismatch repair-deficient colorectal cancer. *Front Oncol.* 2023;13:1223915. doi:10.3389/fonc.2023.1223915.

Supplementary Table 1. PCR primers used for fragment size analysis (n = 30)[a]

Target	Forward Primer	Reverse Primer	Amplicon Size (bp)
TGFBR2 K128	5'-GCTGCTTCTCCAAAGTGCAT-3'	5'-GCACAGATCTCAGGTCCCAC-3'	152
AIM2 T342	5'-TCTGATTGAAGAGGCTGTAT-3'	5'-ATAGTGACTGCAAACAGTGGA-3'	182
ASTE1 R657	5'-GCCTCACTATGTTCCTCTAAGT-3'	5'-CTGAGGCTAAGCAACTTTATG-3'	232
TAF1B N66	5'-TGGTCTTTTTCTCTTTCCTG-3'	5'-AATTCCATTCCATGTCCTG-3'	191
ACVR2A K437	5'-TTGGGAATAGGTGACAGAGT-3'	5'-GCTAACTGGATAACTTACAGCA-3'	250
MSH3 K383	5'-TCTACCAGCTATCTTCTGTGC-3'	5'-TACAAGTACCAAGTAGCCATGA-3'	278
PTEN K267	5'-TCCAATTCAGGACCCACA-3'	5'-GAGTAAGCAAAACACCTGC-3'	511
PTEN N323	5'-ATACATTCTTCATACCAGGACC-3'	5'-GTCAAGCAAGTTCTTCATCAG-3'	296
MLH3 E586	5'-CAAATGTTTCTTGGGCAC-3'	5'-TAAAGATGCTACTGAAGTGGG-3'	266
MLH3 N674	5'-AAAGAGGGGGATGTATCAGA-3'	5'-GTGCCCAAGAAACATTTG-3'	241
EXO1 D731	5'-TGGGGACTCTAGGAATCTG-3'	5'-TAAGCATACACCGCTTCTG-3'	165
LIG3 I157	5'-AGTCTGGGGGTGATATGAA-3'	5'-AAGCAGTGTTCTCTACCTTACC-3'	183
[b]ASXL1 G645	5'-CTTCTCTGAATGGTGTGTTATG-3'	5'-CAGTAGTTGTGTTCGCTGTAGA-3'	548
[b]BCAS3 G891	5'-ACAGACTAACTGAAAGAATCGG-3'	5'-ATGTCACTGTCCTCCTTGAC-3'	463
[b]BEND5 K351	5'-GATTTTCCAGCTAAGGGTATC-3'	5'-CATGAATCTCTTACAGTTGTGC-3'	616
[b]FBXL3 L295	5'-CCACCACACATCTTCACAA-3'	5'-TAGTTTTGTGGTAAGGGAACTG-3'	624
[b]LARP7 T405	5'-GAGGAACTCCCTTAGCTTATTT-3'	5'-ATTCCTACAAACTTGCTCAGTC-3'	625
[b]MARCKS K155	5'-TCCTCGACTTCTTCGCCCAAG-3'	5'-AGGAGAAGCCGCTCAGCTTGA-3'	121
[b]MYH11 P1933	5'-AAATCCAAGCTACACACACAG-3'	5'-GCCTGTAATCCCAGCTACTC-3'	608
[b]PPP3CA N77	5'-GGGTAAGGTATTTGAAGACTCA-3'	5'-GGATGGATATTTGGTGGTTA-3'	385
[b]SYNJ2 P456	5'-CTCAAGTCTGAGAAATACCCAT-3'	5'-ACGTGTCCTGGATTTACAAC-3'	533
[b]WNT11 T181	5'-CTGAGGATGAGGATGGTG-3'	5'-CTTCGTGTATGCGCTGTC-3'	412
[b]TCF7L2 K462	5'-GGCTTAGATCTGGGCACTGTG-3'	5'-ACGGCTGTGCAATTGATGAG-3'	529
[b]TTK R853	5'-GCTTAAGGCCAAGCTTCGTG-3'	5'-GCTATCCACCCACTATTCCAAGAG-3'	536
[b]CASP5 T68	5'-GATAACACTGCATGGGCCTTG-3'	5'-GCATTATTCCCAATGCCACAG-3'	593
[b]SLC35F5 C248	5'-TCCCAACCTGTGTGTGTTTG-3'	5'-CCCCAAAAAGTCTCGTGTGA-3'	452
[b]SLC22A9 K335	5'-CTTTATTCGGTCCCCATCATGG-3'	5'-TCTCTGCGTTGGTGTCCAGTG-3'	665
[b]LTN1 N582	5'-CAAGAGGTTTGGCTTGGACA-3'	5'-TTCTCAAAGACCCAGGATTGC-3'	609
[b]RNF43 G659	5'-TGCACAGTTGCATCCTGG-3'	5'-CTCTCTAACCCACAGTGCC-3'	197
[b]RNF43 R117	5'-CTTAGGAAACATGGGGACA-3'	5'-GTGCAATGCCAGTGATGA-3'	200

[a] PCR Primers for fragment size analysis and amplification prior to Sanger sequencing per target. Nomenclature denotes targeted gene and amino acid position. Primers run for fragment analysis had a FAM tag added to the 5' end of the forward primer and no modification to the reverse primer. Primers used for amplification prior to Sanger sequencing had an M13FW tag (GTAAAACGACGGCCAGT) added to the 5' end of the forward primer and an M13RV tag (GGAAACAGCTATGACCATG) added to the 5' end of the reverse primers. M13FW and M13RV primers were then used for Sanger sequencing.
[b] Targets only tested for Sanger sequencing.

Supplementary Table 2. Thermal cycling conditions per PCR fragment size analysis and Sanger sequencing primer pair

Target(s)	PCR Conditions
TGFBR2 K128	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 60 °C for 1 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
AIM2 T342, TAF1B N66	95 °C for 10 min; then 40 cycles of 95 °C for 30 s, 58 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
ASTE1 R657	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 58 °C for 1 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
ACVR2A K437, MSH3 K383, RNF43 G659	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 58 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
MARCKS K155	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 65 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
PTEN K267, PTEN N323, MLH3 E586, MLH3 N674, EXO1 D731, LIG3 I157, RNF43 R117, ASXL1 G645, BCAS3 G891, BEND5 K351, FBXL3 L295, LARP7 T405, PPP3CA N77, SYNJ2 P456	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 60 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
CASP5 T68, LTN1 N582, SLC22A9 K335, SLC35F5 C248, TCF7L2 K462, TTK R853	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 65 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
WNT11 T181	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 70 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.
MYH11 P1933	95 °C for 10 min; then 40 cycles of 95 °C for 15 s, 68 °C for 2 min; followed by a final extension of 72 °C for 7 min, then holding at 4 °C.

Supplementary Table 3. PCR conditions for GSP1 and GSP2 amplifications of all panels using the ArcherDX LiquidPlex Library Prep Kit

Targets	PCR Conditions
GSP1 PCR	95 °C for 3 min; then 15 cycles of 95 °C for 30 s, 65 °C for 5 min; followed by a final extension of 72 °C for 3 min, then holding at 4 °C.
GSP2 PCR	95 °C for 3 min; then 18 cycles of 95 °C for 30 s, 65 °C for 5 min; followed by a final extension of 72 °C for 3 min, then holding at 4 °C.

GSP1 = gene-specific primers 1; GSP2 = gene-specific primers 2

Supplementary Table 4. Targets utilized in Archer NGS sequencing (n = 168)

ABCC5 p.L1090[c,d]	CDC7 p.N31[c,d]	HIAT1 p.L269[c,d]	MYH11 p.P1940[b,d]	RPL22 p.K15[c,d]	TGFBR2 p.K128[a,d]
ACVR2A p.K437[a,d]	CEBPZ p.D477[c,d]	HMMR p.K667[c]	MYO1A p.K1033[c,d]	RYR2 p.F4739[c]	THAP5 p.K99[c,d]
ADAMTS18 p.F695[c]	CEP162 p.F330[c]	HNRNPL p.P204[c]	NBEAL1 p.E136[c,d]	SEC63 p.K535[c,d]	TM9SF3 p.S72[c,d]
ADD3 p.K701[c,d]	CEP290 p.I556[c,d]	ICA1 p.N203[c,d]	NDUFC2 p.F69[c,d]	SETD1B p.H8[c,d]	TMA16 p.V16[c,d]
ADNP p.K1016[c,d]	CKAP2 p.K606[c,d]	JPH4 p.A502[c,d]	NOL4L p.Y377[c,d]	SGOL1 p.M325[c]	TMEM60 p.A78[c,d]
AIM2 p.T343[a,d]	COBLL1 p.L928[c,d]	KCNMA1 p.K661[c,d]	OR51E2 p.F156[c]	SLAMF1 p.S277[c,d]	TNKS2 p.N622[c,d]
AKAP7 p.K57[c,d]	DDX60L p.N1168[c,d]	KCTD16 p.A384[c,d]	PARP14 p.N1322[c,d]	SLC22A9 p.K335[b,d]	TRAPPC8 p.F679[c,d]
ARSJ p.K579[c,d]	DNA2 p.S779[c,d]	KDM5A p.G1200[c,d]	PDS5B p.K1318[c]	SLC35F5 p.C248[b,d]	TTC3 p.E801[c,d]
ARV1 p.K206[c,d]	DNAH12 p.G93[c]	KIAA1024 p.S702[c,d]	PHACTR4 p.S62[c,d]	SLC35G2 p.R66[c,d]	TTK p.R854[b,d]
ASH1L p.T2895[c]	DOCK3 p.P1852[c,d]	KIAA2018 p.I147[c]	PHF2 p.K492[c]	SLC3A2 p.K300[c]	TTLL10 p.V249[c,d]
ASTE1 p.R657, p.R632[a,d]	DPAGT1 p.F233[c,d]	KIAA2026 p.A730[c]	PLEKHA6 p.V328[c,d]	SLC9A4 p.K798[c]	TVP23A p.F140[c]
ASXL1 p.G645[b,d]	DYNC1I2 p.R57[c,d]	KMT2C p.K2797[c,d]	PPP2R3C p.S23[c]	SMAP1 p.K172[c,d]	UBR5 p.E2121[c,d]
ATAD2 p.E119[c,d]	EBPL p.F172[c,d]	KNOP1 p.I218[c,d]	PPP3CA p.N77[b]	SPAG9 p.R535[c,d]	UPF3A p.E267[c,d]
ATR p.S779[c]	EIF2B3 p.A151[c,d]	LARP4B p.T163[c,d]	PRDM2 p.V1490[c,d]	SPEF2 p.K886[c]	USP35 p.T655[c,d]
BAX p.E41[c,d]	EIF3J p.E61[c,d]	LARP7 p.T405, p.T412[b,d]	PRR11 p.E23[c,d]	SPINK5 p.K823[c,d]	USP40 p.K1148[c]
BCAS3 p.G891[b,d]	EML6 p.F1152[c,d]	LIG3 p.I157[a]	PRRG1 p.P134[c,d]	SRCAP p.P1878[c]	UVRAG p.S237[c,d]
BEND5 p.K351[b]	EPHB2 p.K1020[c,d]	LMAN1 p.E305[c,d]	PRRT2 p.R217[c,d]	SREK1IP1 p.R91[c,d]	VCP p.N616[c,d]
BMPR2 p.N583[c,d]	EXO1 p.D731[a]	LRRIQ3 p.K244[c]	PTEN p.K267[a,d]	SRPRA p.K142[c,d]	VEPH1 p.N554[c]
BRD3 p.P24[c,d]	FAM214A p.N681[c,d]	LTN1 p.N536[b,d]	PTEN p.N323[a]	STAMBPL1 p.K405[c,d]	VEZT p.S26[c,d]
CASP5 p.T68[b]	FAM60A p.N73[c]	MAPRE3 p.C182[c]	QKI p.K134[c]	STAU2 p.N192[c,d]	WDR55 p.K341[c,d]
CCDC168 p.F7017[c,d]	FBXL3 p.L295[b,d]	MARCKS p.K155[b]	RAD50 p.K722[c,d]	SVIL p.M1553[c,d]	WDTC1 p.E290[c,d]
CCDC18 p.V259[c]	FCHO2 p.N15[c]	MDN1 p.K1989[c,d]	RBM27 p.K816[c,d]	SYCP2 p.N1024[c,d]	WNT11 p.T181[b]
CCDC181 p.K421[c]	FHOD3 p.S336[c,d]	MECOM p.G614[c]	RBM45 p.A384[c]	SYNJ2 p.P1113[b,d]	XPOT p.F126[c]
CCDC28A p.N26[c,d]	FXR1 p.N39[c]	MIS18BP1 p.S690[c]	RFC3 p.I82[c,d]	TAF1B p.N66[a,d]	XYLT2 p.G529[c,d]
CCDC43 p.R216[c,d]	GBP3 p.T585[c,d]	MLH3 p.E586[a]	RGS12 p.K1178[c,d]	TBC1D23 p.K649[c,d]	ZBTB20 p.P619[c,d]
CCDC73 p.E448[c]	GLTSCR1 p.P937[c]	MLH3 p.N674[a,d]	RGS22 p.K1150[c,d]	TCERG1 p.K974[c]	ZDBF2 p.K1728[c,d]
CCDC73 p.N638[c,d]	GRB14 p.K297[c,d]	MSH3 p.K383[a,d]	RNF43 G659[b,d]	TCF7L2 p.468[b,d]	ZNF106 p.S438[c,d]
CD3G p.K71[c,d]	GRIK2 p.N849[c,d]	MVK p.A141[c,d]	RNF43 R117[b,d]	TEAD2 p.H167[c,d]	ZNF365 p.K399[c,d]

[a] Present in 12-target panel
[b] Present in 18-target supplemental panel
[c] Present in 138-target supplemental panel
[d] Present in the final 122-target panel

Supplementary Table 5. Summary of FSM detection in supernatant DNA of MSI-H cells via PCR fragment size analysis[a]

Targets	cMNRs	HCT116	LoVo	KM12	HCT15	RKO
ACVR2A K437	A8	m1	m1	m1	m1	m1
AIM2 T343	A10	m1	m1	wt	wt	m1, p1
ASTE1 R657	A11	m1, m2	m1, wt	m1, m2	m1, p1	m2
ASXL1 G645	G8	p1, wt	m1, wt	wt	—	—
BCAS3 G891	A9	m1, wt	wt	wt	—	—
BEND5 K351	A9	wt	wt	wt	wt	wt
CASP5 T68	A10	m1, wt	m1, wt	wt	m1, wt	wt
EXO1 D731T	A7	wt	m1, wt	wt	wt	wt
FBXL3 L295	T9	m1, wt	wt	m1, wt	—	—
LARP7 T405	A8	wt	m1, wt	wt	—	—
LIG3 I157S	A8	wt	wt	wt	wt	wt
LTN1 N582	A11	m1, m2	m1, m2	m3	wt	m2, m3
MARCKS K155	A11	m1	m1	m1, wt	p1, wt	m1, wt
MLH3 E586N	A9	wt	wt	wt	wt	m2, wt
MLH3 N674I	A8	wt	wt	m1, wt	wt	wt
MSH3 K383	A8	m1	wt	wt	wt	m1
MYH11 P1933	C8	wt	m1, p1	m1, wt	—	—
PPP3CA N77	A7	wt	wt	wt	wt	wt
PTEN K267R	A6	wt	wt	m1, wt	wt	wt
PTEN N323M	A6	wt	wt	wt	wt	wt
RNF43 G659	G7	wt	wt	m1	m1, wt	m1
RNF43 R117	C6	m1	wt	wt	wt	wt
SLC22A9 K335	A11	m2, wt	m2, wt	m1, m2	m1, wt	m1, m2
SLC35F5 C248	T10	m1	m1, wt	wt	wt	m1, wt
SYNJ2 P456	C8	m1, wt	m1, wt	wt	—	—
TAF1B N66	A11	m3, wt	m1, wt	m2, wt	wt	m2, wt
TCF7L2 K462	A9	wt	m1, wt	Wt	wt	Wt
TGFBR2 K128	A10	m1	m1, m2	m1, wt	m1, wt	m2, wt
TTK R853	A9	m1, wt	m1, wt	m1, wt	wt	m2, wt
WNT11 T81	A7	wt	wt	wt	wt	wt

[a] Two MSS cell lines (Colo205 and HT29) were also analyzed. None of these loci had FSMs. m = minus (deletion); p = plus (insertion); wt = wild type; — = not tested.

Supplementary Table 6. Number of FSMs detected and percent detection rate in a DNA spike-in assay[a]

Mutant DNA (No. of FSMs detected in cellular DNA)	% mutant DNA spike-in[b], No. (%)					
	5%	1%	0.5%	0.25%	0.10%	0.05%
HCT116 (17)	—	12 (71%)	8 (47%)	6 (35%)	3 (18%)	1 (6%)
KM12 (13)	12 (92%)	12 (92%)	10 (77%)	7 (54%)	3 (23%)	—
LoVo (16)	14 (88%)	12 (75%)	10 (63%)	8 (50%)	4 (25%)	—

[a] A panel of 30-target and the Archer NextSeq platform were used.

[b] MSI-H mutant supernatant DNA was spiked into MSS wt supernatant DNA from HT29. — = not tested.

Supplementary Table 7. The lowest spike-in samples in which FSMs were detected[a]

Target (n = 30)	% Spike-in
ACVR2A K437	0.10%
AIM2 T343	0.25%
ASTE1 R657	0.25%
ASXL1 G645	0.10%
BCAS3 G891	1%
BEND5 K351	—
CASP5 T68	5%
EXO1 D731T	—
FBXL3 L295	0.25%
LARP7 T405	0.25%
LIG3 I157S	—
LTN1 N582	0.10%
MARCKS K155	5%
MLH3 E586N	—
MLH3 N674I	0.25%
MSH3 K383	0.25%
MYH11 P1933	5%
PPP3CA N77	—
PTEN K267R	0.50%
PTEN N323M	—
RNF43 G659	0.10%
RNF43 R117	0.10%
SLC22A9 K335	0.10%
SLC35F5 C248	1%
SYNJ2 P456	0.25%
TAF1B N66	0.10%
TCF7L2 K462	0.25%
TGFBR2 K128	0.25%
TTK R853	0.50%
WNT11 T81	—

[a] Different amount of mutant supernatant DNA from MSI-H cell lines (HCT116, KM12, or LoVo) was serially spiked into WT supernatant DNA from MSS cell line (HT29). — = wt in these three MSI-H cell lines without FSMs detected in spike-in samples.

Supplementary Table 8. Summary of FSM detection in cfDNA from healthy participants

Samples, No. (%)[a]	FSMs detected per sample, No.
1 (2.3)	6
2 (4.5)	5
2 (4.5)	4
7 (15.9)	3
8 (18.2)	2
13 (29.5)	1
11 (25.0)	0

[a] Excluded pooled samples in this analysis

Supplementary Table 9. Number of healthy cfDNA samples with FSMs detected

Target	Samples detected, No. (%)
ADD3 K701	10 (12.0)
USP35 T655	9 (10.8)
PRDM2 V1490	7 (8.4)
FBXL3 L295	6 (7.2)
EBPL F172	5 (6.0)
SETD1B H8	5 (6.0)
AIM2 T343	3 (3.6)
DNA2 S779	3 (3.6)
NDUFC2 F69	3 (3.6)
PRRG1 P134	3 (3.6)
PRRT2 R217	3 (3.6)
ZNF365 K399	3 (3.6)

Supplementary Table 10. Depth and DAF of several targets in matched tumor and cfDNA samples

Patient ID	Target	Ref/Alt Allele	Type	HRUN[a]	Genomic Location (h38)	Tissue type	Depth[b]	DAF[c]	ND DAF Outlier P Value[d]
P30	*XYLT2*	AC/A	DEL	7	chr17:48433966	cfDNA	783	0.0059	0.037897313
						Tumor	104	0.3514	1.11E-16
	SLC22A9	CA/C	DEL	11	chr11:63149670	cfDNA	280	0.12	0.04542599
						Tumor	156	0.4906	1.11E-16
P31	*CCDC168*	CA/C	DEL	9	chr13:103381995	cfDNA	668	0.0303	0.003180953
						Tumor	775	0.0397	0.000180041
	CCDC28A	CA/C	DEL	8	chr6:139097329	cfDNA	344	0.027	0.030572352
						Tumor	709	0.0411	7.98E-05
P34	*GRB14*	TA/T	DEL	10	chr2:165365365	cfDNA	268	0.05	0.020837106
						Tumor	1901	0.0274	0.002731727
	PRRT2	GC/G	DEL	9	chr16:29825015	cfDNA	183	0.0317	0.028691499
						Tumor	628	0.0244	0.000173672
	CCDC43	CT/C	DEL	9	chr17:42756252	cfDNA	584	0.0162	0.031274535
						Tumor	5210	0.0106	0.000410382
P5	*ACVR2A*	TA/T	DEL	8	chr2:148683685	cfDNA	535	0.7946	1.11E-16
						Tumor	4105	0.3262	1.11E-16
	ASTE1	CT/C	DEL	11	chr3:130733046	cfDNA	558	0.0916	2.27E-05
						Tumor	3146	0.2281	1.11E-16
	TGFBR2	GA/G	DEL	10	chr3:30691871	cfDNA	709	0.6201	1.11E-16
						Tumor	4562	0.2096	1.11E-16

[a] HRUN: homopolymer; number of mononucleotides

[b] Depth: The unique molecule sequencing depth using UMI (Unique Molecular Identifier)

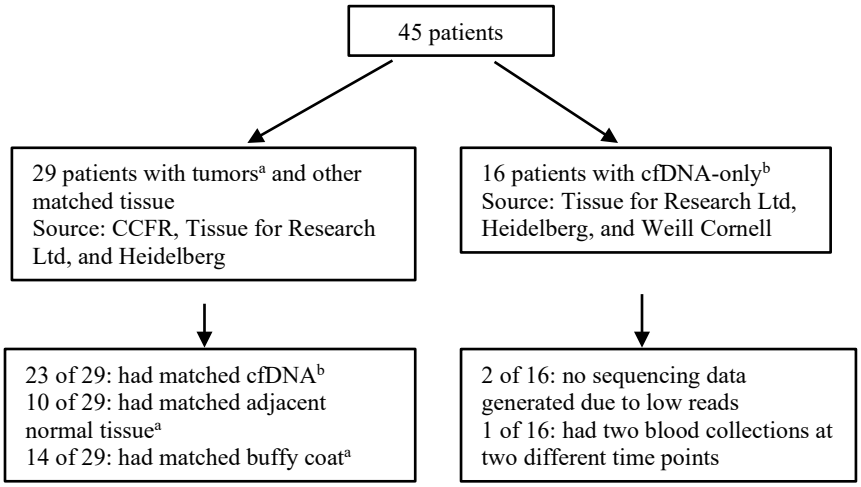
[c] DAF: The allele fraction of the reads from deep (i.e., error-correctable) molecular bins that support the alternative allele (DAO/DDP). DAO is the number of alternate observations (AO) for a variant call from deep (i.e., error-correctable) molecular bins and DDP is the total sequence coverage at this position from deep (i.e., error-correctable) molecular bins.

[d] ND DAF Outlier P Value: Normal Dataset Detectable Allele Fraction) Outlier P-value; The probability this mutation was due to background noise given the provided Normal Dataset and unique molecules with a deep amplicon depth into account. Positive mutation calling was made if ND DAF Outlier $P < 0.05$.

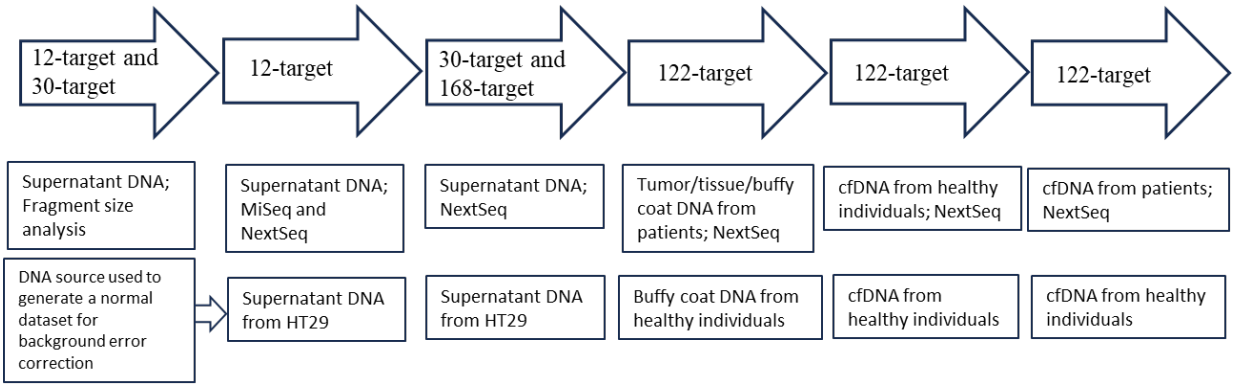
Supplementary Table 11. Number of FSMs detected in matched MSI-H tumor, cfDNA, buffy coat, and adjacent normal tissue with patient clinical information

ID	MMRd[a] or MSI	Gender	AgeDx	Grade	TStage	Targets in tumor, No.	Targets in cfDNA, No.	Targets in buffy coat, No.	Targets in adjacent normal, No.	Targets in both tumor and cfDNA, No.
CCFR										
P30	*MSH2*	M	40	3	4	55	9	1	—	1
P31	*MSH2*	F	44	3	4	35	10	0	—	0
P32	*MSH2*	M	50	2	3	65	—	0	—	—
P33	*MSH2*	M	47	2	2	55	—	0	—	—
P34	*MLH1*	F	73	2	2	57	9	0	—	0
P35	*MLH1*	F	39	3	3	98	2	1	—	0
P36	*MSH2*	M	43	3	3	112	4	0	—	0
P37	*MLH1*	F	42	2	3	109	4	5	—	0
P38	*MSH2*	M	51	—	—	100	—	0	—	—
Tissue for Research Ltd										
P50	MSI-H	F	71	3	pT4a	81	61	—	—	65
P51	MSI-H	F	87	2	pT3	74	56	—	—	38
P52	MSI-H	M	64	2	pT3	71	6	—	—	1
P53	MSI-H	M	62	2	pT4b	35	14[b]	—	—	10
P54	MSI-H	F	71	—	pT1b	38	8[b]	—	—	3
P55	*MLH1, PMS2*	F	61	2	pT4a	63	9	—	4	5
P56	MSI-H	F	53	—	pT3N1a	54	6	0	7	2
P57	MSI-H	F	52	3	pT2a	74	3	—	—	1
P58	*MSH2, MSH6*	F	52	2	pT3	53	—	0	—	—
P59	*PMS2*	M	48	—	pT3	42	—	1	—	—
P60	*MLH1, PMS2*	F	56	3	pT1b	57	—	0	—	—
P61	*MSH2*	F	79	—	—	1	5	0	—	0
P62	*EPCAM*[c]	M	51	4	—	—	50	—	—	—
Heidelberg										
P1	*MLH1*(unclassified variant)	F	46	2	pT3	100	4[b]	—	22	1
P2	*MLH1*	M	28	2	pT3	63	2[b]	—	4[b]	0
P3	*PMS2* VUS	M	72	—	pT3	78	1[b]	—	19	1
P4	—	F	74	3	pT3	70	21[b]	—	6[b]	19
P5	—	M	72	—	—	89	81	—	26	79
P6	*PMS2*	F	44	3	pT4b	53	9	—	9[b]	4
P7	*PMS2* VUS	M	49	—	pT3	52	3[b]	—	17	0
P8	—	M	36	3	pT2	69	2	—	32	1

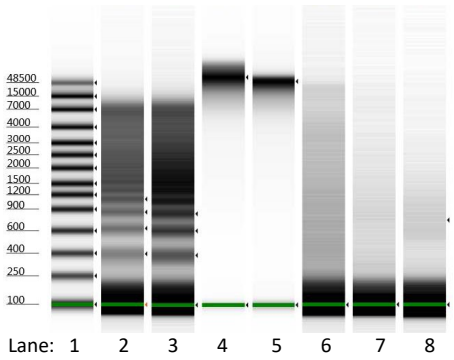
[a] Assessed mainly by immunohistochemical staining of MMR gene products
[b] QC failed due to low unique DNA start sites per gene-specific primers 2 (GSP2) (<50)
[c] DEL/DUP EXB_3'UTRdel
VUS = variant with unknown significance; — = No matched samples available



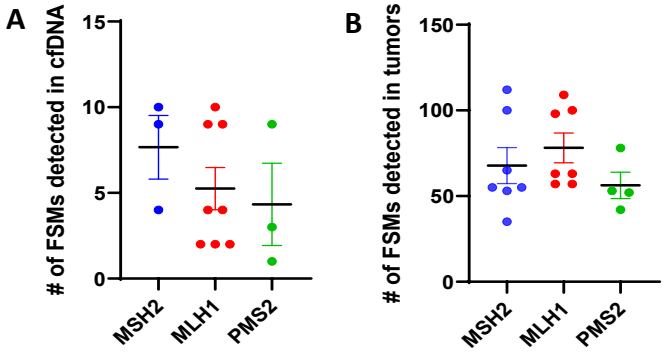
Supplementary Figure 1. Diagram of patient samples used in the study.
[a] Data from tumors, adjacent normal tissue, and buffy coat was reported in Table 4.
[b] Data from cfDNA was reported in Table 6.



Supplementary Figure 2. Flow diagram to show the sources of DNA (including DNA tissue sources used to generate a normal dataset for background error correction for NGS data analysis) and platforms used at the various panel development stages.



Supplementary Figure 3. DNA size analysis by Agilent TapeStation system. Lane 1: DNA size marker; Lane 2 and 3: DNA from LoVo and HCT116 culture supernatants; Lane 4 and 5: cellular DNA from HCT116 and LoVo; and Lanes 6, 7, and 8: cfDNA from plasma_EDTA, plasma_ACD and serum from healthy participants.



Supplementary Figure 4. Number of FSMs detected in **(A)** cfDNA and **(B)** tumors by MMR gene. MMR status was performed by immunohistochemical staining and data was not avialable for some samples. There was no significant difference among groups for either cfDNA or tumors. However, several limitations should be considered when interpreting the data (e.g., small small size, low and different plasma or serum volume used from each sample, different tumor stage and grade, etc.).