Supplementary Materials

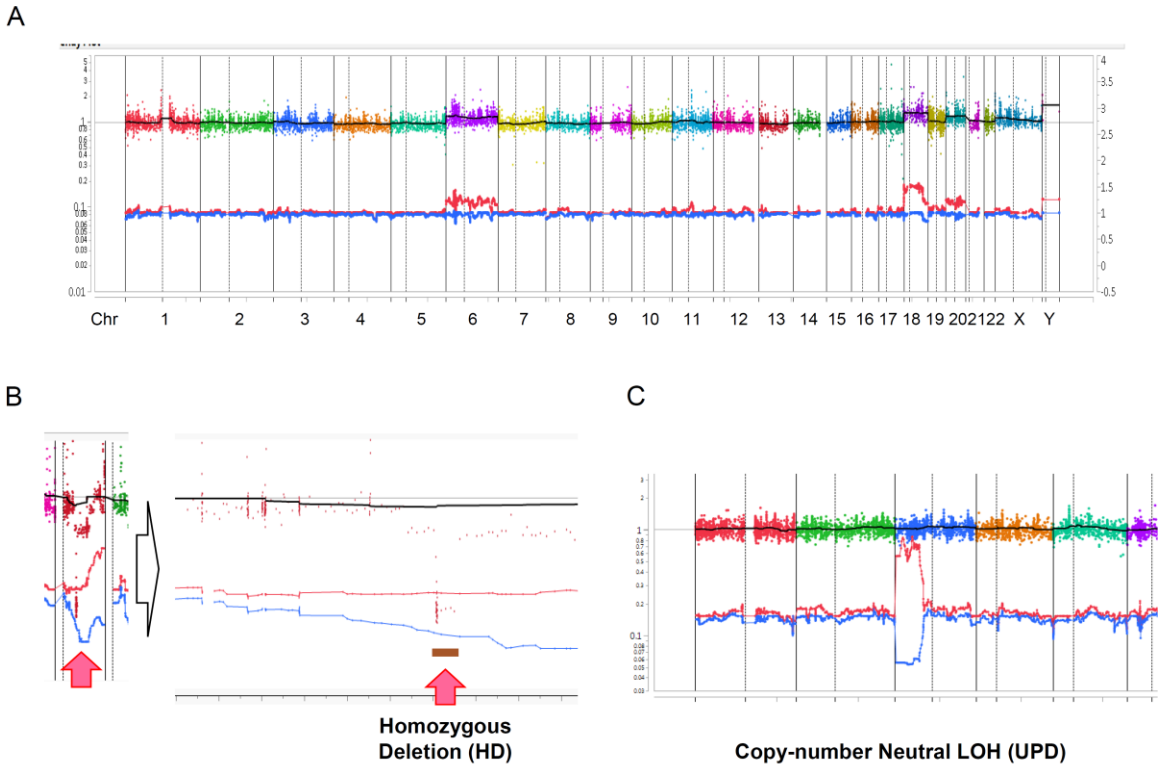

Figure S1. CNA analysis with TOP DNA panel.

(A) Total (upper, left axis) and allelic (lower red and blue lines, right axis) somatic CNA throughout chromosomes. **(B)** An example of focal homozygous deletion detection with TOP DNA panel. **(C)** An example of copy-number neutral loss of heterozygosity (uniparental disomy).

Case #	Fusion_gene	Gene1 (5' gene)	Gene2 (3' gene)	Gene1_chr	Gene2_chr	RefSeq_gene1	RefSeq_gene2	Gene1_last_ observed_exon / Total exon	Gene2_first_ observed_exon / Total exon	Gene_1_ wild-type_ read_#	Gene_2_ wild-type _read_#	Fusion_ read_#	Estimation
#23	*CCDC6-RET*	*CCDC6*	*RET*	chr10	chr10	NM_005436	NM_020630	1/9	12/19	445	4	0	Negative
#23	*CCDC6-RET*	*CCDC6*	*RET*	chr10	chr10	NM_005436	NM_020630	8/9	12/19	1456	4	0	Negative
#23	*CD74-NRG1*	*CD74*	*NRG1*	chr5	chr8	NM_004355	NM_004495	6/8	6/6	29127	0	0	Negative
#23	*CD74-ROS1*	*CD74*	*ROS1*	chr5	chr6	NM_004355	NM_002944	6/8	34/43	29127	701	0	Negative
#23	*CLTC-ROS1*	*CLTC*	*ROS1*	chr17	chr6	NM_004859	NM_002944	31/32	35/43	1985	1467	0	Negative
#23	*EML4-ALK*	*EML4*	*ALK*	chr2	chr2	NM_019063	NM_004304	6/23	20/29	1572	0	0	Negative
#23	*EML4-ALK*	*EML4*	*ALK*	chr2	chr2	NM_019063	NM_004304	13/23	20/29	2780	0	0	Negative
#23	*EML4-ALK*	*EML4*	*ALK*	chr2	chr2	NM_019063	NM_004304	20/23	20/29	555	0	107	Positive
#23	*EML4-ALK*	*EML4*	*ALK*	chr2	chr2	XM_005264267	NM_004304	7/24	20/29	26	0	0	N.D.
#23	*EML4-ALK*	*EML4*	*ALK*	chr2	chr2	XM_005264267	NM_004304	7/24	18/29	26	2	0	N.D.
#23	*ERC1-RET*	*ERC1*	*RET*	chr12	chr10	NM_178039	NM_020630	7/18	12/19	511	4	0	Negative
#23	*ERC1-RET*	*ERC1*	*RET*	chr12	chr10	NM_178039	NM_020630	12/18	12/19	534	4	0	Negative
#23	*ERC1-RET*	*ERC1*	*RET*	chr12	chr10	NM_178039	NM_020630	17/18	12/19	48	4	0	N.D.
#23	*EWSR1-FLI1*	*EWSR1*	*FLI1*	chr22	chr11	NM_005243	NM_002017	7/17	6/9	1094	501	0	Negative
#23	*EZR-ROS1*	*EZR*	*ROS1*	chr6	chr6	NM_003379	NM_002944	9/13	34/43	1392	701	0	Negative
#23	*FUS-CREB3L2*	*FUS*	*CREB3L2*	chr16	chr7	NM_004960	NM_194071	6/15	5/12_not_entire	1157	318	0	Negative
#23	*KIF5B-RET*	*KIF5B*	*RET*	chr10	chr10	NM_004521	NM_020630	15/26	12/19	1007	4	0	Negative
#23	*NAB2-STAT6*	*NAB2*	*STAT6*	chr12	chr12	NM_005967	NM_003153	4/7	5/22	164	631	0	Negative
#23	*NCOA4-RET*	*NCOA4*	*RET*	chr10	chr10	NM_005437	NM_020630	7/10	12/19	881	4	0	Negative
#23	*PAX3-FOXO1*	*PAX3*	*FOXO1*	chr2	chr13	NM_181457	NM_002015	7/8	2/3	0	142	0	Negative
#23	*PAX7-FOXO1*	*PAX7*	*FOXO1*	chr1	chr13	NM_002584	NM_002015	7/8	2/3	60	142	0	Negative
#23	*RET-CCDC6*	*RET*	*CCDC6*	chr10	chr10	NM_020630	NM_005436	11/19	2/9	3	444	0	Negative
#23	*RET-NCOA4*	*RET*	*NCOA4*	chr10	chr10	NM_020630	NM_005437	11/19	8/10	3	879	0	Negative
#23	*SDC4-NRG1*	*SDC4*	*NRG1*	chr20	chr8	NM_002999	NM_004495	4/5	6/6	914	0	0	Negative
#23	*SLC34A2-ROS*	*SLC34A2*	*ROS1*	chr4	chr6	NM_006424	NM_002944	13/13_not_entire	32/43	2467	983	0	Negative
#23	*SS18-SSX1*	*SS18*	*SSX1*	chr18	chrX	NM_005637	NM_005635	9/10	6/8	414	0	0	Negative
#23	*SS18-SSX2*	*SS18*	*SSX2*	chr18	chrX	NM_005637	NM_003147	9/10	6/9	414	0	0	Negative
#23	*STRN-ALK*	*STRN*	*ALK*	chr2	chr2	NM_003162	NM_004304	3/18	20/29	1112	0	0	Negative
#23	*TPM3-ROS1*	*TPM3*	*ROS1*	chr1	chr6	NM_152263	NM_002944	8/10	35/43	4782	1467	0	Negative
#23	*TRIM27-RET*	*TRIM27*	*RET*	chr6	chr10	NM_006510	NM_020630	3/8	12/19	288	4	0	Negative
#23	*TRIM33-RET*	*TRIM33*	*RET*	chr1	chr10	NM_015906	NM_020630	11/20	12/19	576	4	0	Negative

Figure S2. Validation of the junction capture method for each putative fusion gene.

A pipeline was developed to count the wild-type transcripts of both genes involved in fusion transcripts around putative fusion junctions according to the COSMIC database to verify the validity of the junction capture RNA-seq method. A representative result of this validation step for case #23 (*EML4-ALK*-positive lung adenocarcinoma) is partially summarized. The number of sequencing reads aligned to the wild-type transcript of the putative fusion genes are calculated in the column Gene_1 or Gene_2_wild-type_read_#. The existence of fusion genes was estimated using the following criteria: Positive, Fusion_read_# >0; Negative, Gene_1 or Gene_2_wild-type_read_# ≥50 and Fusion_read_#=0; or N.D. (not determined), Gene_1 and Gene_2_wild-type_read_# <50 and Fusion_read_#=0.

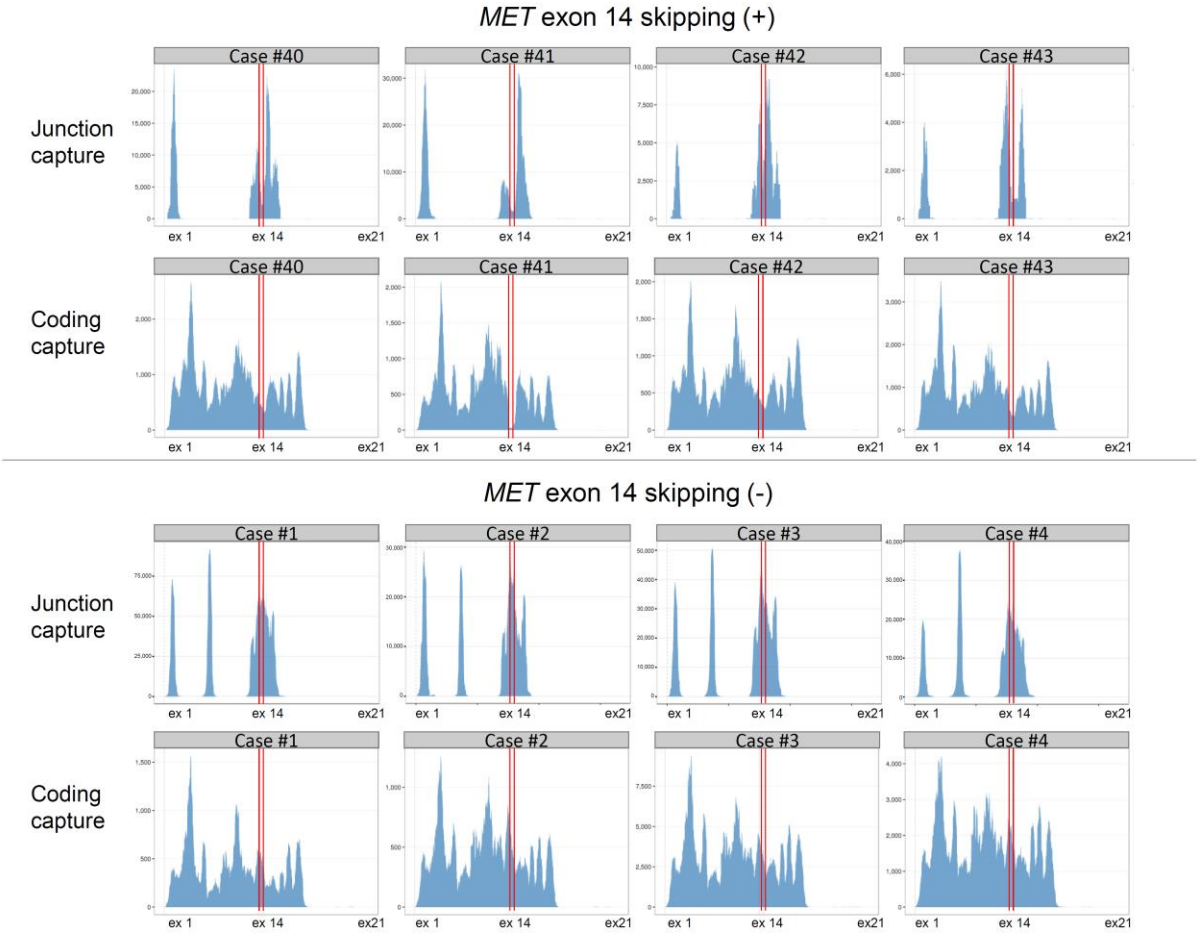

Figure S3. Detection of *MET* exon 14 skipping using the junction capture or coding capture method.

RNA-seq reads of *MET* exon 14 skipping-positive cases (#39-#43) and *MET* exon 14 skipping-negative cases (#1-#4) mapped to virtual MET cDNA constructed on the x-axis that corresponds to the transcript of NM_000245. RNA-seq was performed using the junction capture method or the coding exon capture method (TruSight RNA Pan-Cancer Panel, Illumina) to synthesize cDNA libraries from the RNA extracted from FFPE samples. Regions between the red lines indicate *MET* exon 14.

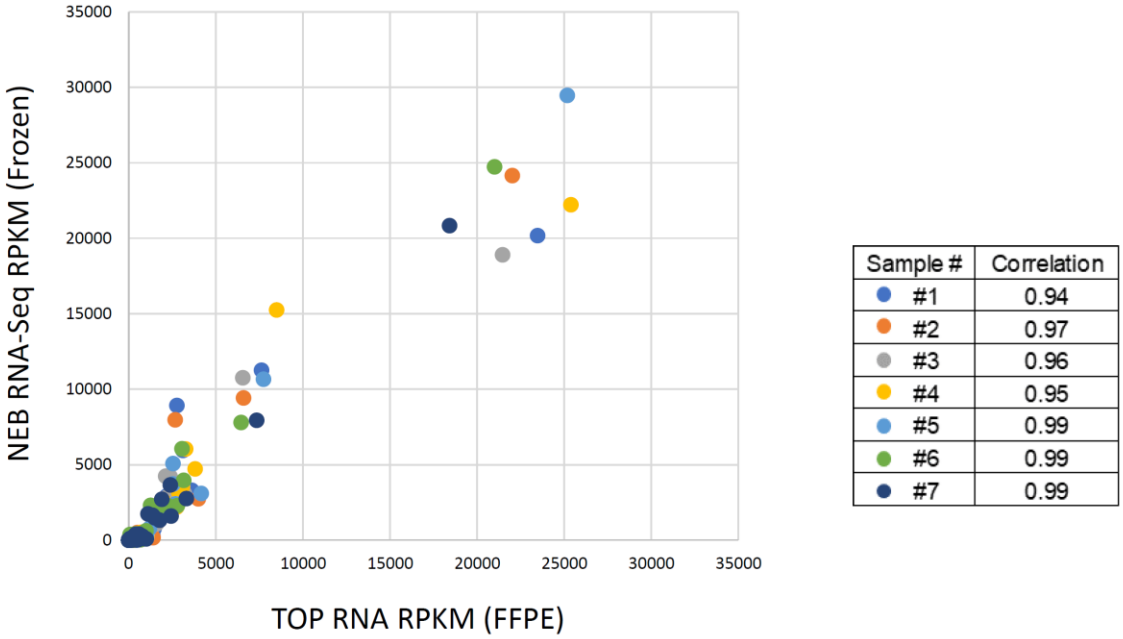

Figure S4. Validation of expression analysis using the TOP RNA panel.

Expression analysis was performed for 7 tumors to compare the performance of the TOP RNA panel using FFPE specimens with that of RNA-seq using frozen specimens and poly(A) mRNA enrichment. The mRNA expression values of 109 genes using FFPE and frozen specimens of identical tumors evaluated by the TOP RNA panel and by RNA-seq using poly(A) mRNA enrichment (NEBNext Ultra Directional RNA Library Prep Kit, NEB), respectively, were quantified, and their correlations were evaluated.

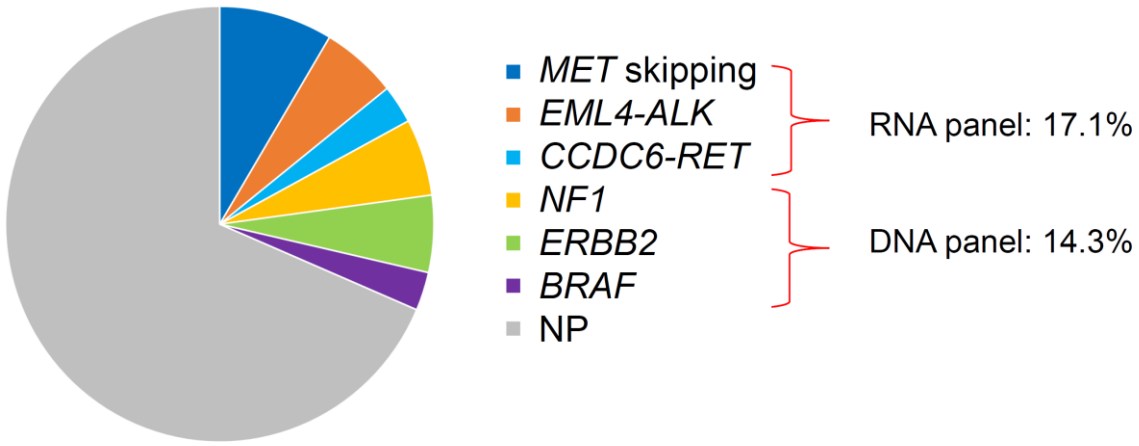

Figure S5. Retrospective TOP sequencing in NSCLC.

Driver oncogenes identified by junction capture RNA-seq (TOP RNA panel V3) and gDNA target sequencing (TOP DNA panel V1) in a cohort of patients with stage II-III NSCLC whose cancers were negative for *KRAS* or *EGFR* mutations.

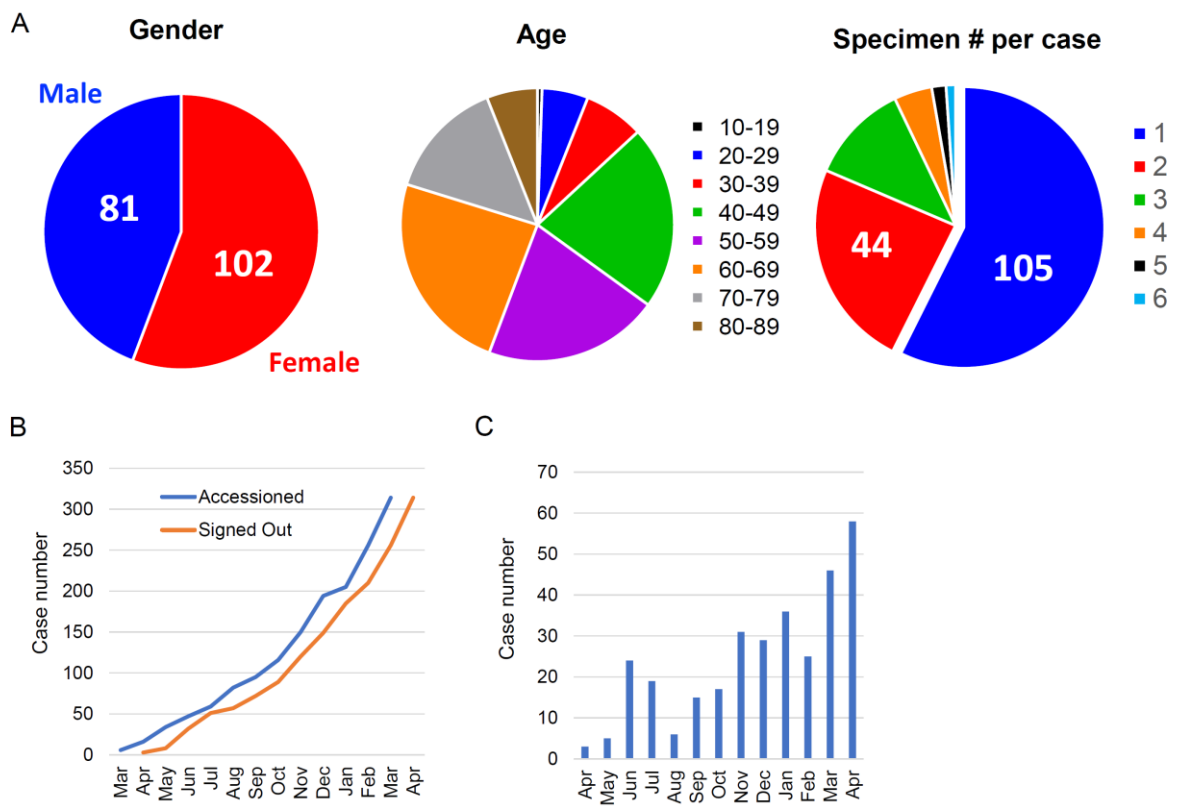

Figure S6. Features of the TOP prospective cohort.

(A) Patient gender and age and the number of specimens submitted for each individual patient are indicated. **(B)** The accrual of samples in the TOP prospective cohort for the duration of this study. The blue line indicates the number of samples that were accessioned into the laboratory, while the orange line indicates the samples that were successfully sequenced with a clinical report issued into the medical record of the patient. **(C)** The number of samples that were successfully sequenced during each month.

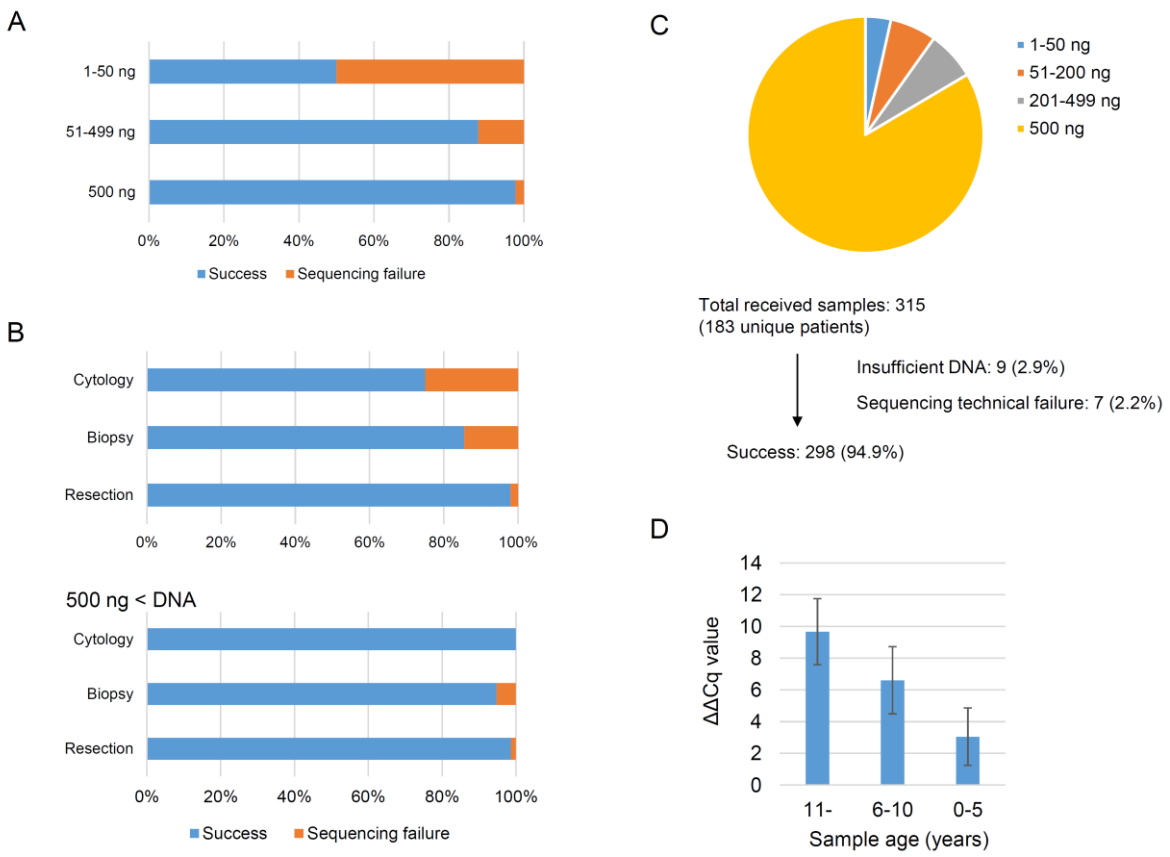

Figure S7. TOP DNA panel sequencing success as a function of specimen characteristics.

(A) Assay performance as a function of genomic DNA used for sequence library preparation. Samples with the optimal DNA input of 500 ng, which constituted 84% of all sequenced samples, achieved the highest success rate (98%), whereas samples with DNA input ranging from 1-50 ng achieved the lowest success rate (50%) but still produced informative results for half of the cases. **(B)** Assay performance as a function of specimen type. Resections had the highest overall success rate (98%), followed by biopsies (86%) and cytology samples (75%). **(C)** Distribution of DNA input across all sequenced samples. **(D)** A DNA quality check was conducted by calculating ΔΔCq values using TaqMan FFPE DNA QC Assay v2 kits. DNA quality was affected by sample age.

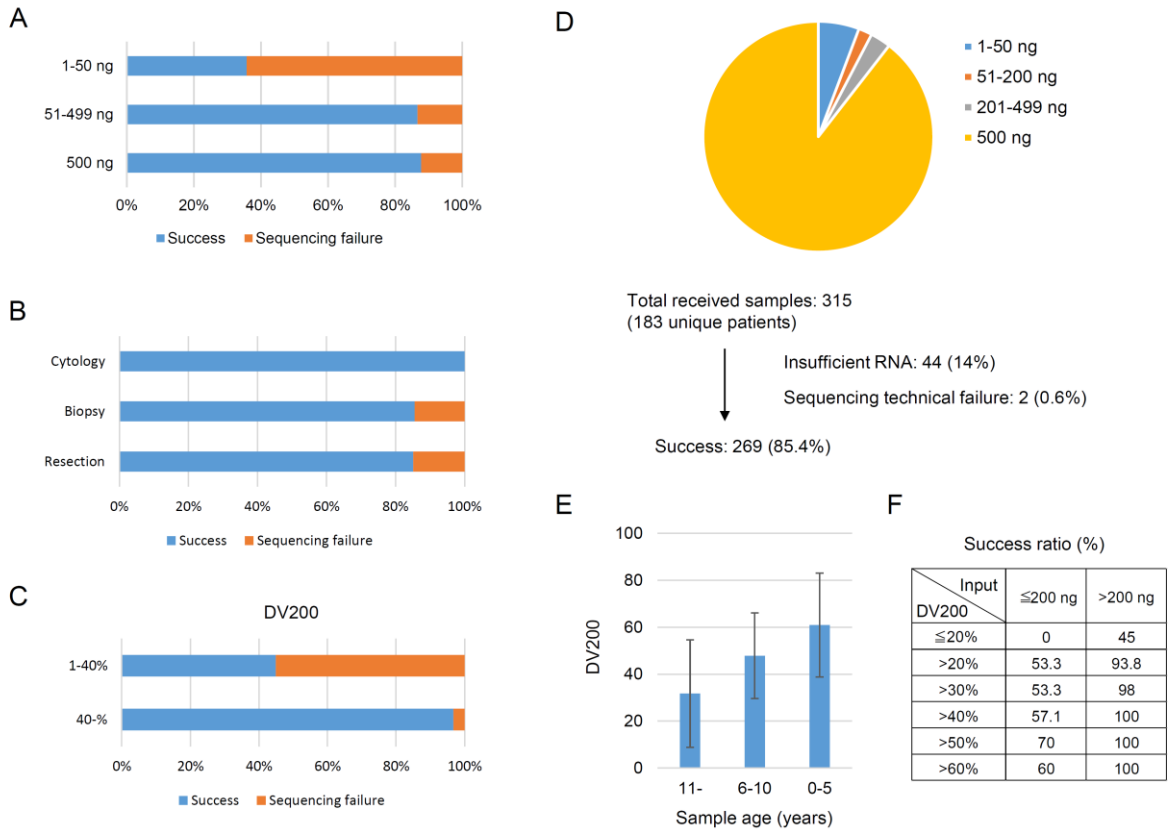

Figure S8. TOP RNA panel sequencing success as a function of specimen characteristics.

(A) Assay performance as a function of RNA used for sequence library preparation. Samples with the optimal RNA input of 500 ng, which constituted 91% of all sequenced samples, achieved the highest success rate (88%), whereas samples with RNA input ranging from 1-50 ng achieved the lowest success rate (36%). **(B)** Assay performance as a function of specimen type. Cytology samples had the highest overall success rate (100%: 4 out of 4), followed by biopsies (86%) and resection samples (85%). **(C)** Assay performance as a function of DV200. An RNA quality check was conducted by calculating DV200 values using the HighSensitivity RNA Screen Tape system. Samples with a DV200 >40% had the best overall success rate of 97%, whereas samples with a DV200 ≤40% had an overall success rate of 45%. **(D)** Distribution of RNA input across all sequenced samples. **(E)** DV200 values, RNA

quality scores, were affected by sample age. **(F)** The TOP RNA panel success ratio was evaluated using the initial RNA input and DV200 values. The success ratio of samples with an RNA input >200 ng and a DV200 >40% was 100%.

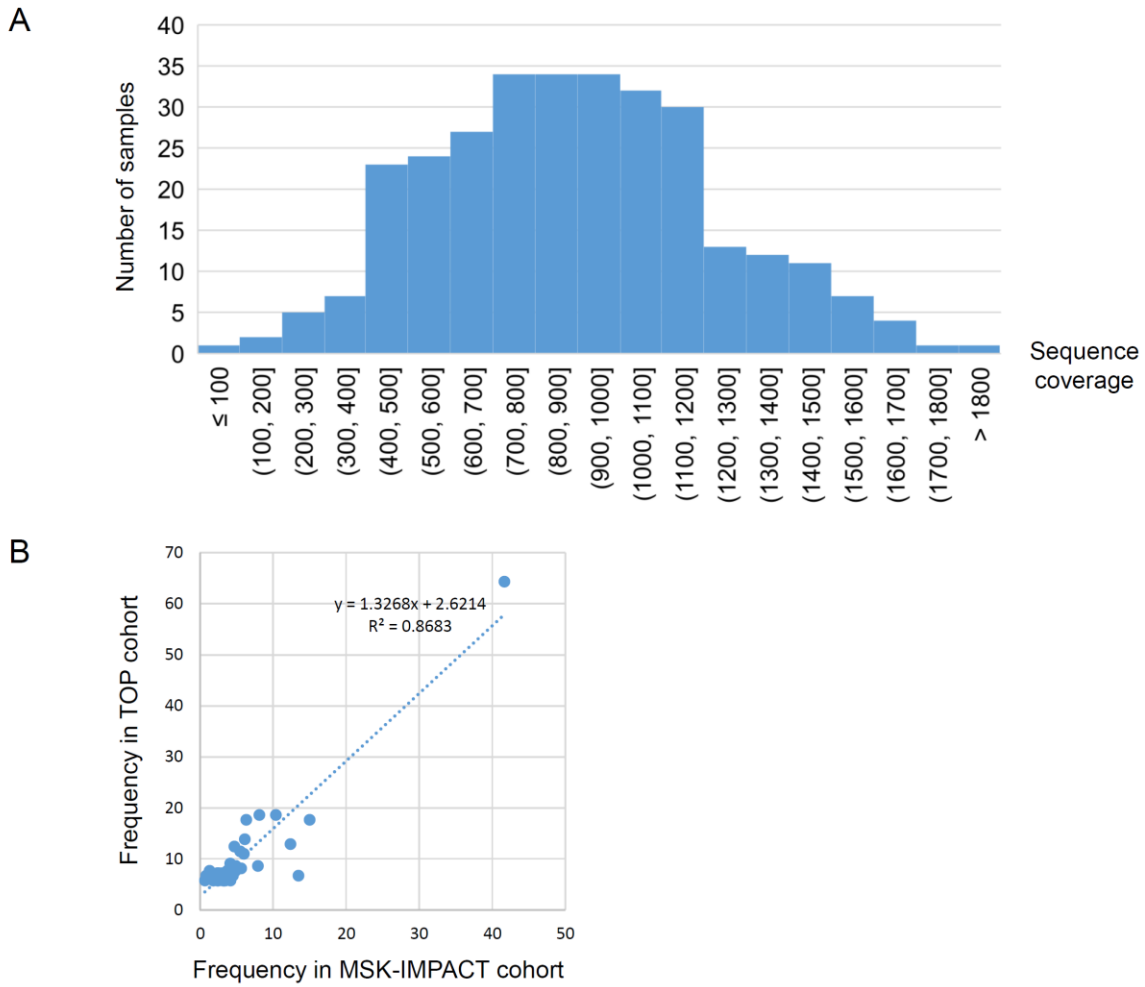

Figure S9. TOP DNA panel sequence coverage and major mutation frequency.

(A) Distribution of mean unique sequence coverage for reported samples sequenced using the TOP DNA panel. **(B)** Frequency of gene alterations in the TOP and MSK-IMPACT cohorts. Genes whose mutation frequency was in the top 20 in the TOP cohort were selected, and the mutation frequencies of those genes were compared in the TOP and MSK-IMPACT cohorts.

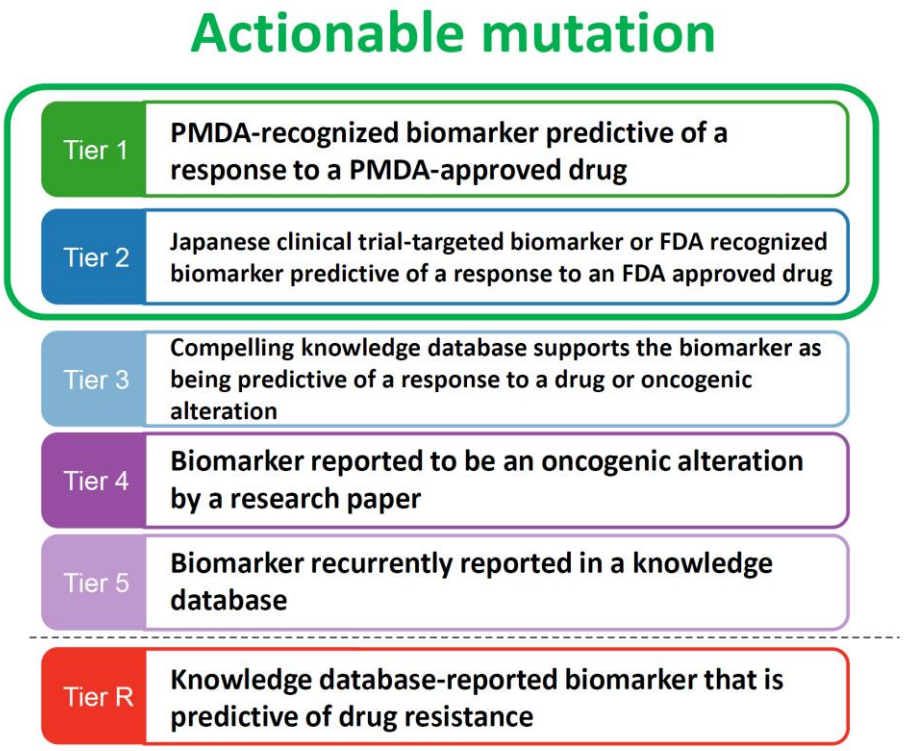

Figure S10. TOP evidence level classification.

The evidence level classification used to annotate gene alterations in TOP testing is indicated.

PMDA: Pharmaceuticals and Medical Devices Agency; FDA: Food and Drug Administration.

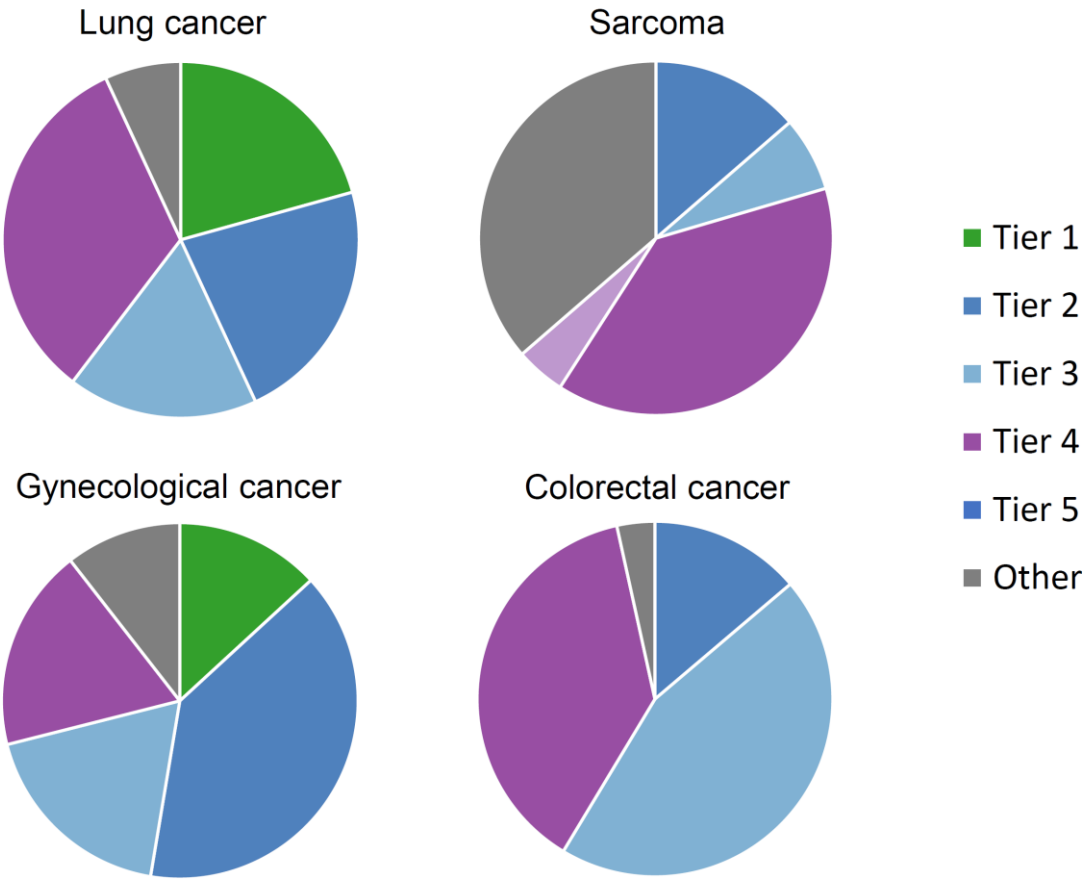

Figure S11. Clinical actionability of somatic alterations revealed by the TOP in several types of cancers.

Alterations were annotated based on their clinical actionability according to TOP classification (**Figure S9**), and the samples were assigned to the level with the most actionable alteration.

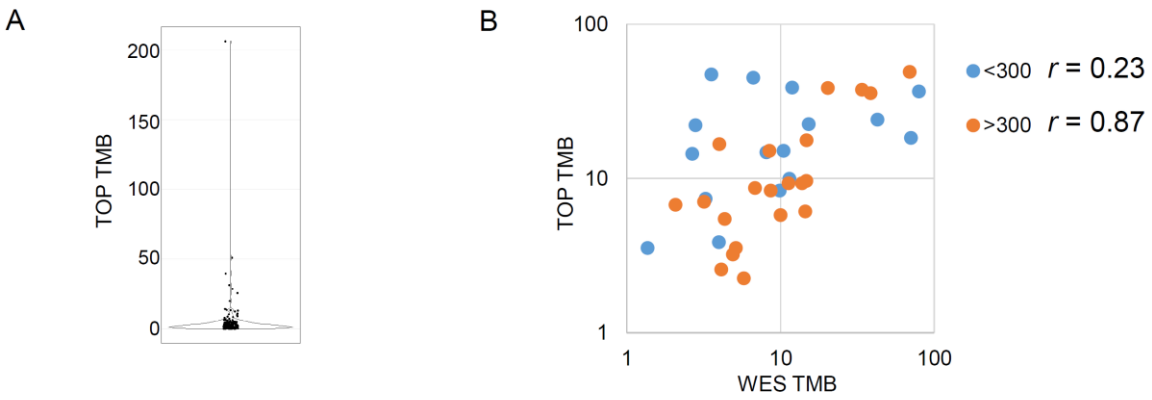

Figure S12. Tumor mutation burden calculation using the TOP DNA panel.

(A) Violin plots show the somatic tumor mutation burden (TMB) distribution, defined as the number of coding mutations per megabase. The width plot indicates the frequency of samples with a given TMB. The distribution of observed mutation rates across all sequenced tumors was used to identify a threshold indicative of high mutation burden: 8.5 mutations/Mb. **(B)** TMB correlation between the TOP and whole exome sequencing. The TOP DNA panel and WES were performed using FFPE and fresh frozen specimens, respectively, for the preparation of DNA libraries from 37 tumors to compare the TMB calculated by the TOP with that calculated by WES.

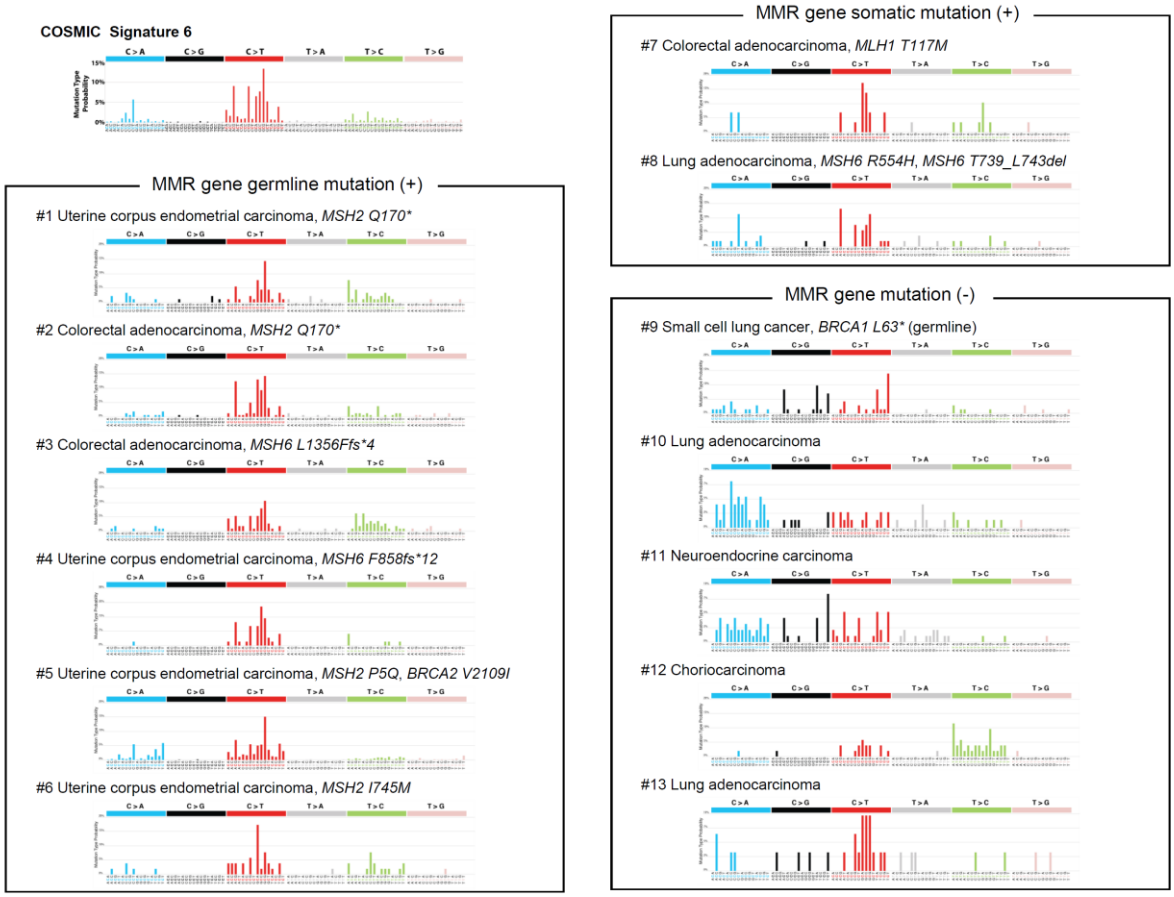

Figure S13. Mutation signatures of samples with high mutation burdens.

Using the pattern and nucleotide context of all observed silent and nonsilent substitutions in the 13 samples (4.1%) with elevated mutation rates (>8.5 mutations/megabase (Mb); **Figure S11A**), mutations in each sample were assigned to constituent mutation signatures from the set of 30 signatures described previously. Using this approach, we identified tumors with mutations in mismatch-repair (MMR) genes such as *MSH2*, *MSH6* or *MLH1* that showed Signature 6, which is associated with defective DNA mismatch repair and microsatellite-unstable tumors. Samples #1 and #2 are from the same patient.

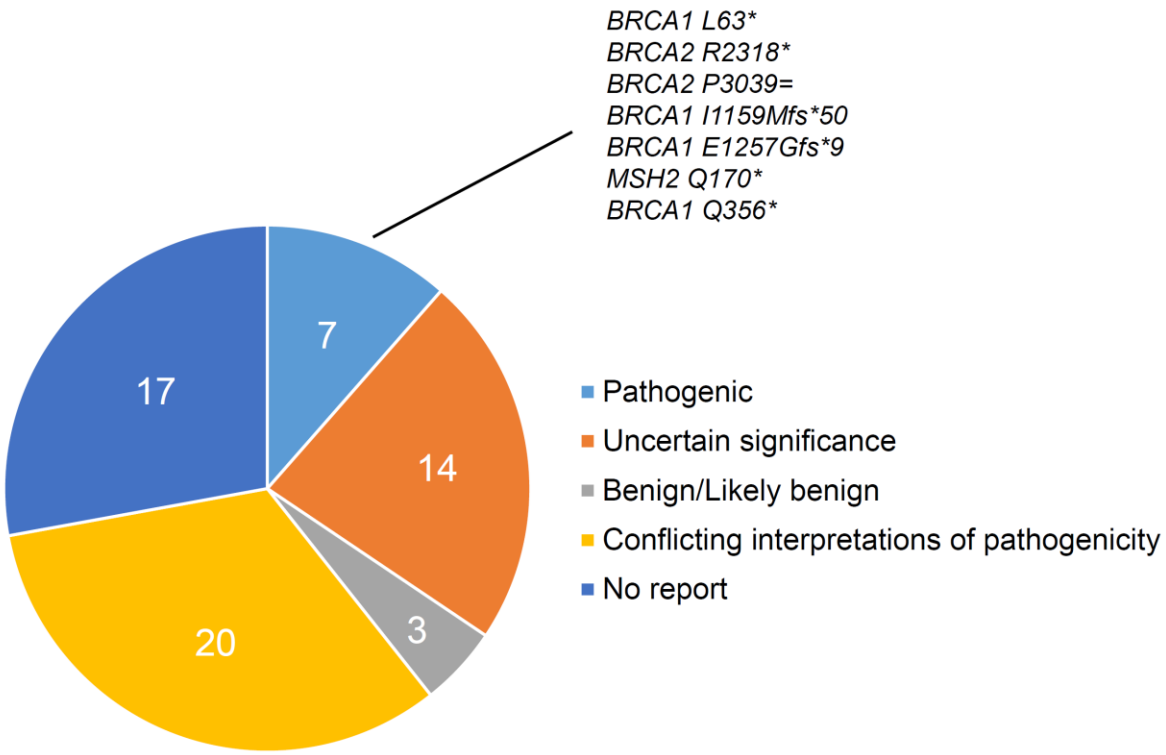

Figure S14. Germline mutations found in the TOP cohort.

Germline mutations found in the TOP prospective cohort and their pathogenicity.

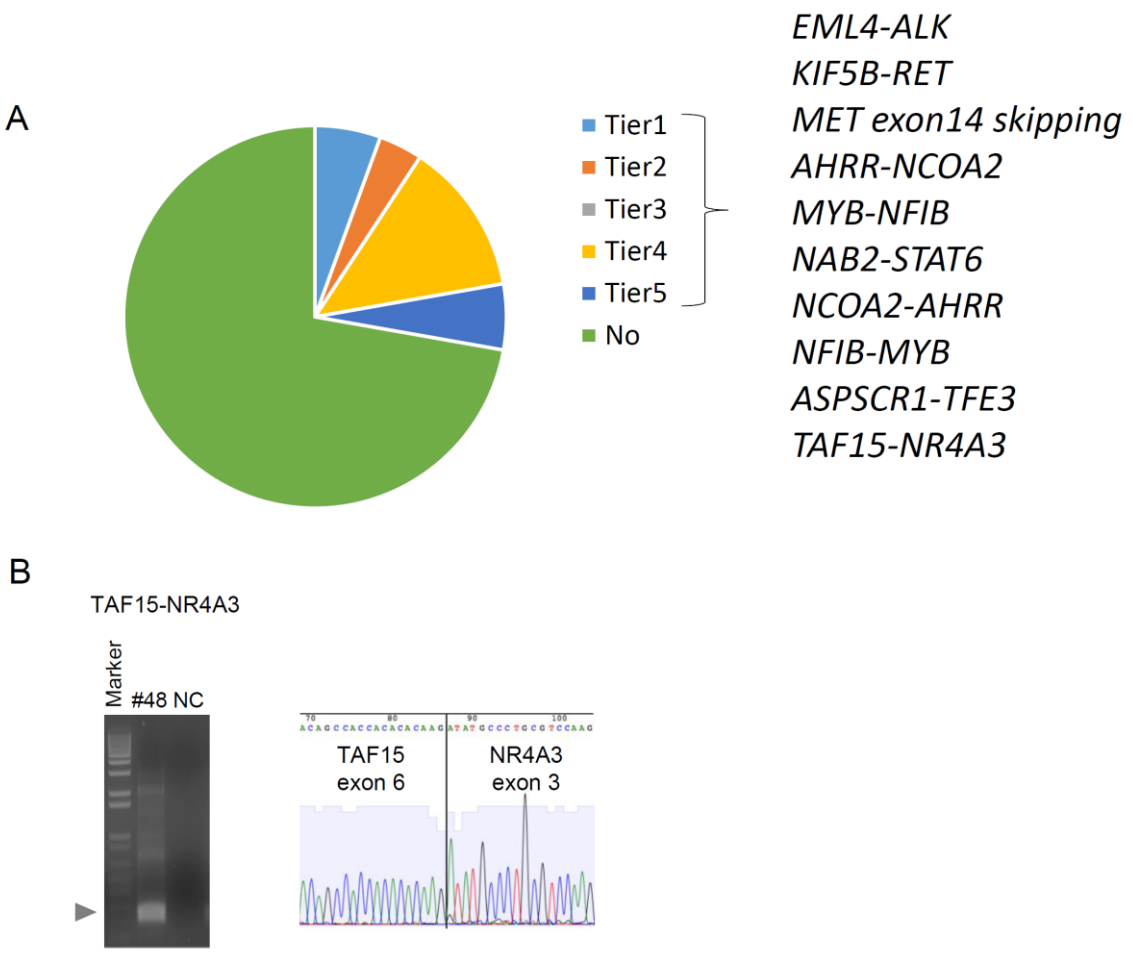

Figure S15. Clinical actionability of transcriptional alterations revealed by the TOP RNA panel.

(A) The clinical actionability of the transcriptional alterations were annotated based on TOP classification (**Figure S9**). **(B)** A fresh frozen sample from case #48 was subjected to RT-PCR with the fusion-RT primer set to detect TAF15-NR4A3 mRNA (left panel). The arrowhead indicates the estimated size of the fusion transcript. The band was extracted from the gel and subjected to Sanger sequencing. The electrophoretogram obtained from the band supported the junction sequence of TAF15-NR4A3 (right panel). Marker, 1-kb DNA ladder; NC, negative control.

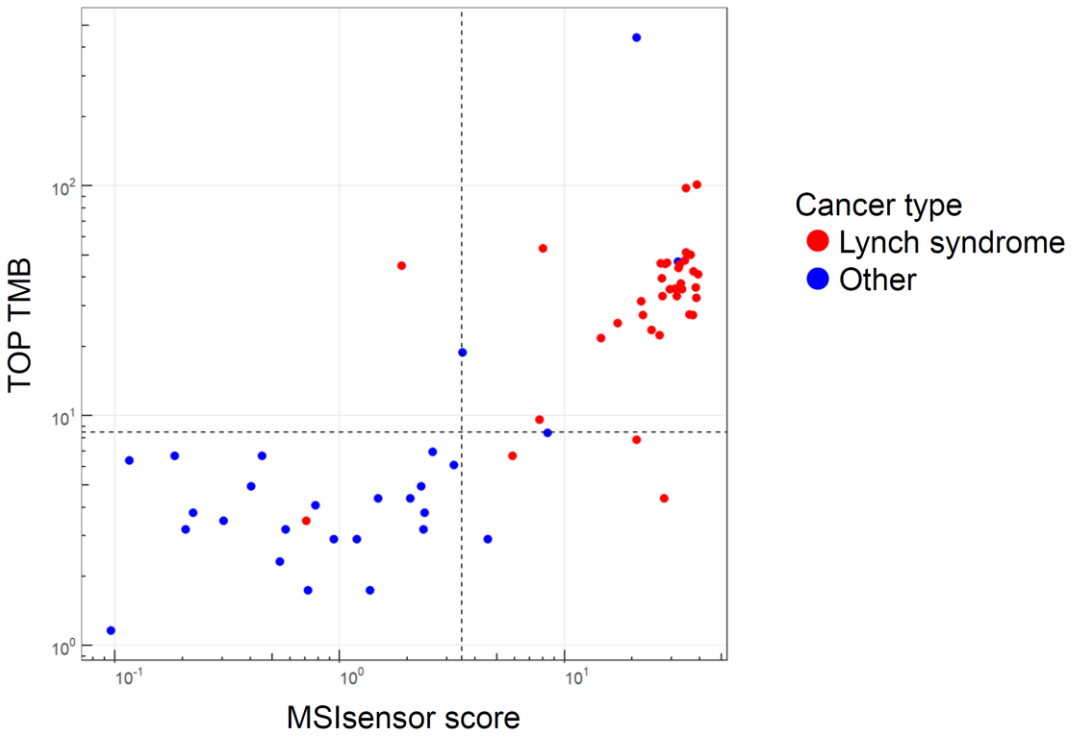

Figure S16. MSIsensor score using TOP DNA sequences.

The MSIsensor scores and TOP TMBs are compared for 36 specimens of Lynch syndrome (red dots) and 27 specimens of other cancer types with unknown MMR statuses (blue dots).