Supplemental Online Content

Hayashi H, Takiguchi Y, Minami H, et al. Site-specific and targeted therapy based on molecular profiling by next-generation sequencing for cancer of unknown primary site: a nonrandomized phase 2 clinical trial. *JAMA Oncol*. Published online October 15, 2020. doi:10.1001/jamaoncol.2020.4643

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods

Molecular Assay Procedures

Macrodissection of FFPE samples was performed to select a cancer region, which was marked by a pathologist after paraffin removal. DNA and RNA were extracted and purified from the samples with the use of an Allprep DNA/RNA FFPE Kit (Qiagen, Valencia, CA). The quality and quantity of the DNA and RNA were verified with a PicoGreen dsDNA Assay Kit (Life Technologies, Foster City, CA) and RiboGreen RNA Assay Kit (Life Technologies). The extracted DNA and RNA were stored at −80°C until analysis.

For DNA sequencing, 20 ng of DNA were subjected to multiplex PCR (polymerase chain reaction) amplification with the use of an Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific, Waltham, MA) and Ion AmpliSeq Cancer Hotspot Panel v2 (Thermo Fisher Scientific). Construction and purification of libraries were performed as described.[1] Pooled libraries were sequenced with an Ion Torrent PGM instrument (Thermo Fisher Scientific) and with the use of an Ion PGM Hi-Q Chef Kit (Thermo Fisher Scientific) and Ion 316 v2 Chip Kit (Thermo Fisher Scientific). DNA sequencing data were accessed through Torrent Suite v.5.0.2 software. Reads were aligned with the hg19 human reference genome, and variants were called with the use of Variant Caller ver 5.0.2. and then processed as described previously.[1-3]

For RNA sequencing, PCR primers were designed with the use of Ion AmpliSeq Designer (Life Technologies). The genes subjected to RNA sequencing are listed in eTable 1. An Ion AmpliSeq RNA Library Kit (Thermo Fisher Scientific) was used for RNA library construction from 10 ng of total RNA. Pooled libraries were sequenced with an Ion Torrent PGM instrument and with the use of an Ion PGM Hi-Q Chef Kit and Ion 316 v2 Chip Kit. Relative expression (raw read number/total read number) of each gene was calculated for normalization.

Data Analysis

On the basis of the gene expression profiles of 130 CUP patients analyzed in our previous,[4] we selected marker genes that were significantly expressed in CUP with predicted tissue of origin based on the classification using Bayesian posteriori probability maximization method.[4]

To predict the tissue of origin, we used a weighted voting algorithm, in which each weight value was calculated as the signal-to-noise ratio.[5] Briefly, the signal-to-noise statistic (S_x) is calculated as $S_x = (\mu_{class0} - \mu_{class1}) / (\sigma_{class0} + \sigma_{class1})$, where μ is the mean value and σ is the standard deviation for each class. In this scheme, the

algorithm can also be used to find the decision boundaries between the class means as $b_x = (\mu_{class0} + \mu_{class1}) /2$ for each gene. To predict the class of a test sample γ, each gene x in the predictive gene set has a vote based on the expression in this sample (g_x^{γ}) and b_x: $V_x=S_x (g_x^{\gamma}-b_x)$ and the final vote for class 0 or 1 is sign ($\sum_x V_x$).

A validation test was performed to estimate the accuracy (true classification rate) of the classification procedure with the gene data about 16 cancer types from TCGA data set. RNASeqV2 data for those cancer types were downloaded from TCGA Data Portal upon request to the Data Coordinating Center (DCC). As a result, the average yielded a value of 85.0% (5,428/6389) as a validated estimate of accuracy as shown in eTable2.

We also used Bayesian inference in order to estimate the tissue of origin based on the gene alteration. The posterior probability was calculated with the following formula based on the mutation rate found in cBioPortal database.

$$P(\text{Ori}|\text{mut}) = \frac{P(\text{mut}|\text{Ori})*P(\text{Ori})}{\sum_{\text{ori}} P(\text{mut}|\text{Ori}) *P(\text{Ori})}$$

P(mut|Ori) was calculated based on the number of patients with corresponding genetic alteration divided by the number of patients analyzed for each cancer type. These numbers were based on studies summarized by cBioPortal (https://www.cbioportal.org/), such as LUAD (n=230, TCGA, Nature 2014), Lung SCC (n=178, TCGA, Nature 2012), COAD (n=74, Genentech, Nature 2012 & n=138, MSKCC, Genome Biol 2014).[6-17] P(Ori) was the number of new cancer cases from Cancer statistics, 2014 (https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21208). P(Ori|mut) was calculated for male and for female, respectively. The example about the calculation of prior probability for ERBB2 was shown in eTable3.

When the gene alteration was found for those genes, including *KRAS*, *EGFR*, *HER2* and *KIT*, based on the DNA sequencing result, the prior probability was applied for the patient together with the estimation based on the gene expression based on RNA-Seq profiles. Tissue from the study patients was classified with the rule, and the resulting prediction (the class with the highest posteriori probability) was reported to the treating physician.

References

1. Takeda M, Sakai K, Terashima M, et al. Clinical application of amplicon-based next-generation sequencing to therapeutic decision making in lung cancer. *Ann Oncol.* 2015;26(12):2477-2482.
2. Sakai K, Takeda H, Nishijima N, et al. Targeted DNA and RNA sequencing of fine-needle biopsy FFPE specimens in patients with unresectable hepatocellular carcinoma treated with sorafenib. *Oncotarget.* 2015;6(25):21636-21644.
3. Takeda M, Sakai K, Okamoto K, et al. Genome sequencing for nonsmall-cell lung cancer identifies a basis for nintedanib sensitivity. *Ann Oncol.* 2016;27(4):748-750.
4. Hayashi H, Kurata T, Takiguchi Y, et al. Randomized Phase II Trial Comparing Site-Specific Treatment Based on Gene Expression Profiling With Carboplatin and Paclitaxel for Patients With Cancer of Unknown Primary Site. *J Clin Oncol.* 2019;37(7):570-579.
5. Yanagisawa K, Tomida S, Shimada Y, Yatabe Y, Mitsudomi T, Takahashi T. A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. *J Natl Cancer Inst.* 2007;99(11):858-867.
6. Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature.* 2014;507(7492):315-322.
7. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61-70.
8. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487(7407):330-337.
9. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature.* 2013;499(7456):43-49.
10. Schulze K, Imbeaud S, Letouze E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet.* 2015;47(5):505-511.
11. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014;511(7511):543-550.
12. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. *Nature.* 2011;474(7353):609-615.
13. Biankin AV, Waddell N, Kassahn KS, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature.* 2012;491(7424):399-405.
14. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell.* 2015;163(4):1011-1025.

15. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014;513(7517):202-209.
16. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015;517(7536):576-582.
17. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519-525.

eTable1. Gene List for RNA Sequencing

GENE_ID=DHRS2	GENE_ID=ORM1	GENE_ID=ADAP1	GENE_ID=FGF19	GENE_ID=OMD	GENE_ID=TG
GENE_ID=KRT17	GENE_ID=NAPSA	GENE_ID=ADCY2	GENE_ID=FGF3	GENE_ID=OR51E2	GENE_ID=TIMM8A
GENE_ID=KRT7	GENE_ID=NKX2-1	GENE_ID=ADIPOQ	GENE_ID=FGF9	GENE_ID=PAEP	GENE_ID=TIMP4
GENE_ID=S100A2	GENE_ID=SCGB1A1	GENE_ID=ADRA1B	GENE_ID=FKBP11	GENE_ID=PAPOLG	GENE_ID=TMPRSS3
GENE_ID=SDC1	GENE_ID=SFTPA2	GENE_ID=AKAP6	GENE_ID=FOLH1	GENE_ID=PART1	GENE_ID=TPO
GENE_ID=UPK3A	GENE_ID=SFTPB	GENE_ID=AMD1	GENE_ID=FOLH1B	GENE_ID=PAX4	GENE_ID=TRAFD1
GENE_ID=EFHD1	GENE_ID=SFTPC	GENE_ID=ARHGAP6	GENE_ID=G6PC	GENE_ID=PCNXL2	GENE_ID=TRIM9
GENE_ID=ERBB2	GENE_ID=LRMP	GENE_ID=ASTN2	GENE_ID=GAL3ST1	GENE_ID=PDLIM5	GENE_ID=TRPM8
GENE_ID=ESR1	GENE_ID=MS4A1	GENE_ID=ATP10B	GENE_ID=GAP43	GENE_ID=PDZK1	GENE_ID=TRPV6
GENE_ID=GATA3	GENE_ID=PAX8	GENE_ID=ATP2C1	GENE_ID=GATA2	GENE_ID=PGM3	GENE_ID=TSHR
GENE_ID=MUCL1	GENE_ID=RGS13	GENE_ID=ATP6V1G2	GENE_ID=GFAP	GENE_ID=PHF1	GENE_ID=TTR
GENE_ID=PIP	GENE_ID=EMX2	GENE_ID=BAALC	GENE_ID=GJB5	GENE_ID=PLEK2	GENE_ID=TTTY14
GENE_ID=SERPINB4	GENE_ID=FBXO21	GENE_ID=BAAT	GENE_ID=GPM6A	GENE_ID=PLIN1	GENE_ID=TXNL4A
GENE_ID=CDH17	GENE_ID=FOLR1	GENE_ID=BCAN	GENE_ID=GPM6B	GENE_ID=PLP1	GENE_ID=VIL1
GENE_ID=CDX1	GENE_ID=KLHDC8A	GENE_ID=BHMT	GENE_ID=GPR135	GENE_ID=PMF1	GENE_ID=ZC3HAV1
GENE_ID=CDX2	GENE_ID=KLK8	GENE_ID=BPIFA1	GENE_ID=GPX2	GENE_ID=PMP2	GENE_ID=ZNF365
GENE_ID=CEACAM5	GENE_ID=MEIS1	GENE_ID=C10orf57	GENE_ID=GRIA3	GENE_ID=PRR15L	GENE_ID=ZNF395
GENE_ID=EPS8L3	GENE_ID=WT1	GENE_ID=C14orf105	GENE_ID=HBD	GENE_ID=PRR4	
GENE_ID=GUCA2A	GENE_ID=CELA3A	GENE_ID=C19orf21	GENE_ID=HNF1B	GENE_ID=RAB3B	
GENE_ID=KRT20	GENE_ID=CLPS	GENE_ID=C1orf61	GENE_ID=HOXB13	GENE_ID=RDH11	
GENE_ID=LGALS4	GENE_ID=CPA1	GENE_ID=C2	GENE_ID=HS3ST3A1	GENE_ID=RNF43	
GENE_ID=PIGR	GENE_ID=CPA2	GENE_ID=C7orf63	GENE_ID=HSD3B2	GENE_ID=RUFY3	
GENE_ID=SLC26A3	GENE_ID=CPB1	GENE_ID=CALB2	GENE_ID=IBSP	GENE_ID=RUNX1	
GENE_ID=TMPRSS4	GENE_ID=CTRB2	GENE_ID=CDH6	GENE_ID=KANSL1L	GENE_ID=RUNX2	
GENE_ID=CAPN14	GENE_ID=GCG	GENE_ID=CDK18	GENE_ID=KCND2	GENE_ID=SATB2	
GENE_ID=GJB6	GENE_ID=GP2	GENE_ID=CDK19	GENE_ID=KCNJ16	GENE_ID=SCN1A	
GENE_ID=KRT4	GENE_ID=PRSS1	GENE_ID=CEACAM1	GENE_ID=KCNN2	GENE_ID=SERPINB3	
GENE_ID=LY6K	GENE_ID=REG1A	GENE_ID=CHGB	GENE_ID=LPHN3	GENE_ID=SFRP2	
GENE_ID=DAPK1	GENE_ID=REG1B	GENE_ID=CKM	GENE_ID=LPP	GENE_ID=SFTPD	
GENE_ID=LRP2	GENE_ID=CELA3B	GENE_ID=COL10A1	GENE_ID=MALT1	GENE_ID=SIM2	
GENE_ID=MYBPC1	GENE_ID=ACPP	GENE_ID=CTNND2	GENE_ID=MAP4K2	GENE_ID=SLC16A4	
GENE_ID=HEPH	GENE_ID=KLK2	GENE_ID=CUL3	GENE_ID=MAU2	GENE_ID=SLC17A1	
GENE_ID=MT3	GENE_ID=KLK3	GENE_ID=CWH43	GENE_ID=MB	GENE_ID=SLC17A3	
GENE_ID=TKTL1	GENE_ID=COMP	GENE_ID=DCLK2	GENE_ID=MIA3	GENE_ID=SLC22A2	
GENE_ID=KRT14	GENE_ID=FBN1	GENE_ID=DLK1	GENE_ID=MLC1	GENE_ID=SLC25A37	
GENE_ID=KRT6B	GENE_ID=DSG3	GENE_ID=DLL3	GENE_ID=MMP13	GENE_ID=SLC30A4	
GENE_ID=S100A7	GENE_ID=GBP6	GENE_ID=DNAJB12	GENE_ID=MRPS18C	GENE_ID=SLC3A1	
GENE_ID=SPRR1A	GENE_ID=KRT5	GENE_ID=DTNA	GENE_ID=MS4A3	GENE_ID=SLC44A1	
GENE_ID=CDH16	GENE_ID=KRT6A	GENE_ID=EGLN3	GENE_ID=MSMB	GENE_ID=SORD	
GENE_ID=FXYD2	GENE_ID=SPRR1B	GENE_ID=ELAC2	GENE_ID=MYH2	GENE_ID=SOX2	
GENE_ID=NAT8	GENE_ID=SPRR3	GENE_ID=ENO3	GENE_ID=MYO1D	GENE_ID=SPAG8	
GENE_ID=APOC3	GENE_ID=CTSE	GENE_ID=ENPEP	GENE_ID=NCAN	GENE_ID=SPDEF	
GENE_ID=C6	GENE_ID=GKN1	GENE_ID=ENPP1	GENE_ID=NOVA1	GENE_ID=SPRY4	
GENE_ID=CRP	GENE_ID=LIPF	GENE_ID=ESM1	GENE_ID=NPDC1	GENE_ID=ST8SIA3	
GENE_ID=FGA	GENE_ID=PGC	GENE_ID=FAM120A	GENE_ID=NPY	GENE_ID=STMN4	
GENE_ID=FGB	GENE_ID=SPINK1	GENE_ID=FAM66D	GENE_ID=NPY2R	GENE_ID=TCL6	
GENE_ID=FGG	GENE_ID=TFF1	GENE_ID=FBXL8	GENE_ID=NR1H4	GENE_ID=TEF	
GENE_ID=HP	GENE_ID=TFF2	GENE_ID=FERMT1	GENE_ID=OAZ3	GENE_ID=TFF3	

eTable2: The results of validation study about tumor classification.

ID	BLCA	BRCA	COAD	KDNY	LIHC	LUAD	OV	PAAD	PRAD	DLBCL	Germ	SARC	SC_all	CESC	HNSC	LUSC
BLCA	**217**	42	0	4	0	1	0	2	0	0	2	12	9	7	2	0
BRCA	3	**964**	0	0	0	1	0	0	7	0	0	3	16	7	9	0
COAD	11	3	**496**	0	1	2	0	3	0	0	26	4	10	9	0	1
KDNY	0	0	0	**854**	2	0	1	0	0	0	2	1	15	14	0	1
LIHC	0	8	0	11	**311**	19	0	0	0	0	9	1	2	0	0	2
LUAD	4	1	0	0	0	**503**	1	0	0	0	2	3	141	6	0	135
OV	3	6	0	2	1	1	**197**	1	0	0	2	46	19	15	1	3
PAAD	0	14	2	0	1	1	1	**103**	0	0	6	17	3	2	0	1
PRAD	1	0	0	0	0	0	0	0	**461**	0	0	0	0	0	0	0
DLBCL	8	32	0	1	1	2	0	4	0	**46**	5	8	6	1	1	4
Germ	14	9	4	1	2	13	60	1	0	1	**97**	15	60	12	16	32
SARC	7	42	1	3	0	0	0	3	0	0	0	**93**	5	2	3	0
SC_all	79	39	1	0	0	3	0	0	0	0	3	1	**1086**	173	507	406
NA	0	1	0	1	0	0	0	0	0	0	2	11	0	0	0	0
total	347	1161	504	877	319	546	260	117	468	47	156	215	1372	248	539	585
ID	BLCA	BRCA	COAD	KDNY	LIHC	LUAD	OV	PAAD	PRAD	DLBCL	Germ	SARC	SC_all	CESC	HNSC	LUSC
CESC	66	7	0	0	0	1	0	1	0	0	0	0	259	**143**	89	27
HNSC	30	43	1	0	0	0	0	0	0	0	0	1	501	29	**414**	58
LUSC	1	0	0	0	0	11	0	0	0	0	0	0	277	0	4	**273**
NA	0	8	0	0	0	0	0	0	0	0	0	0	1	1	0	0

Each column corresponds to the cancer type being analyzed, and each row represents the number of cases for each cancer type as estimated cancer of origin.
In the case of BRCA, 964 out of 1161 samples (83%) were correctly predicted as samples with breast cancer.

eTable3: The example of prior probability for *ERBB2* gene alteration

				①	②:male		③:female			
ID	Origin	Sample no.	Mut. no.	$P(\mathrm{mut}\mid\mathrm{ori})^{1)}$	$P_m(\mathrm{ori})^{2)}$	①×②	$P_f(\mathrm{ori})^{3)}$	①×③	$P_m(\mathrm{ori}\mid\mathrm{mut})$	$P_f(\mathrm{ori}\mid\mathrm{mut})$
BLCA	Bladder	127	4	3.1%	56,390	1,776	18,300	576	15%	2%
BRCA	Breast	482	65	13.5%	2,360	318	232,670	31,377	3%	**83%**
COAD	Colon	212	5	2.4%	48,450	1,143	48,380	1,141	10%	3%
KIRC	Kidney	415	0	0.3%	39,140	116	24,780	74	1%	0%
LIHC	Liver	196	1	0.5%	24,600	126	8,590	44	1%	0%
LUAD	Lung	230	6	2.6%	46,400	1,210	43,284	1,129	10%	3%
OV	Ovarian	316	4	1.3%	NA	NA	21,980	278	NA	1%
PAAD	Panc	90	1	1.1%	23,530	261	22,890	254	2%	1%
PRAD	Prostate	302	3	1.0%	233,000	2,315	NA	NA	20%	NA
STAD	Stomach	287	36	12.5%	13,730	1,722	8,490	1,065	15%	3%
LMH	Lymphoma	48	0	0.0%	38,270	0	32,530	0	0%	0%
GERM	Germ cell	156	0	0.0%	8,820	0			0%	
SARC	Sarcoma	1484	0	0.0%	6,550	0	5,470	0	0%	0%
CE	Cervical	36	2	5.6%	NA	NA	12,360	687	NA	2%
ESO	Eso	126	15	11.9%	14,660	1,745	3,510	418	15%	1%
HN	Head_Neck	302	9	3.0%		0		0	0%	0%
LUSC	Lung	178	5	2.8%	29,000	815	27,053	760	7%	2%
	Total				584,900	11,548	510,287	37,803	100%	100%

eTable4. Details of Poststudy Cancer Therapy

	Site-Specific Therapy (*n* = 97)
Subsequent radiotherapy, n (%)	20 (20.6%)
Subsequent systemic therapy,[a] n (%)	56 (57.7%)[b]
Chemotherapy	
Platinum agent	26 (26.8%)
Taxane	25 (25.8%)[c]
Gemcitabine	14 (14.4%)
5-Fluorouracil	11 (11.3%)
Anthracycline	5 (5.2%)
Topoisomerase inhibitor	11 (11.3%)
Vinca alkaloid	2 (2.1%)
Other[c]	9 (9.3%)
Molecularly targeted agents	
VEGF inhibitor	8 (8.2%)
EGFR-TKI	1 (1.0%)
Endocrine therapy	1 (1.0%)
Immunotherapy	8 (8.2%)
Experimental therapy	4 (4.1%)

[a]Patients may have received more than one type of subsequent therapy.

[b]Number of patients who received at least one line of systemic therapy.

[c]Includes eribulin, trifluridine/tipiracil, cyclophosphamide, pemetrexed, and trabectedin.

Abbreviations: VEGF, vascular endothelial growth factor; EGFR-TKI, epidermal growth factor receptor–tyrosine kinase inhibitor.

eTable 5. Adverse Events of Any Grade in at Least 10% of Patients (n = 97)

	All Grades		Grade ≥3	
Adverse Event	**No.**	**%**	**No.**	**%**
Nonhematologic				
AST increased	48	49.5	6	6.2
Anorexia	46	47.4	7	7.2
ALT increased	43	44.3	6	6.2
Nausea	41	42.3	6	6.2
Fatigue	36	37.1	5	5.2
Constipation	29	29.9	0	0
Creatinine increased	25	25.8	1	1.0
Sensory neuropathy	24	24.7	1	1.0
Fever	21	21.6	1	1.0
Diarrhea	19	19.6	4	4.1
Alopecia	19	19.6	1	1.0
Hematologic				
WBC decreased	53	54.6	21	21.6
Neutrophil decreased	58	59.8	27	27.8
Thrombocytopenia	63	64.9	11	11.3
Anemia	62	63.9	15	15.5

Adverse events are defined according to MedDRA preferred terms. Abbreviations: AST, aspartate aminotransferase; ALT, alanine aminotransferase; WBC, white blood cell.

eFigure1. Patient Disposition