Case Report

Dual Analyses of DNA-seq and RNA-seq Identified Two Independent Transcripts Caused by a Single Nucleotide Variant of *MUTYH* c.934-2A>G in Advanced Colon Cancer

Hideki YAMAMOTO[*1], *Takao HINOI*[*2], *Tomokazu WAKATSUKI*[*3], *Izumi MIYABE*[*4], *Gou YAMAMOTO*[*5], *Miho KAKUTA*[*6], *Junichi ZAITSU*[*7], *Kazuya KURAOKA*[*8], *Kiyomi TANIYAMA*[*9] and *Kiwamu AKAGI*[*10]

We present a case showing a significant advantage in the combined analysis of DNA-seq and RNA-seq, by which two truncated transcripts of *MUTYH* was identified in a case of suspected hereditary colorectal cancer of a woman in her forties. We introduced next-generation sequencing (NGS) -based dual analysis of DNA and RNA as a routine protocol for the germline testing in the mismatch repair (MMR) genes and the suspected splice variants of the listed genes in our multi-gene panel. While no evidence of Lynch syndrome was identified, NGS analysis revealed a substitution of a nucleotide c.145G>C in the transactivation domain of *TP53*, leading to p.Asp49His (D49H), and a heterozygous single nucleotide alteration of a c.934-2A>G in the splice acceptor site of intron 10 of *MUTYH*. RNA-seq additionally uncovered two independent splice variants of *MUTYH* derived from a c.934-2A>G. One variant was a truncated form due to a premature stop codon in the middle of intron 10; another was three-amino-acid shorter form created by the read-through transcript of intron 10 until the additional nine bases located in an alternative 3' splice site selection in exon 11. This case indicates an advantage in the combined analysis of RNA-seq with DNA-seq as a comprehensive germline analysis to identify unknown variants that could not be detected by DNA-seq only. The combined analysis of DNA and RNA might have a potential impact on the identification of etiology and/or development of hereditary tumor that carries splicing variants. 【Case Report】

［Rinsho Byori **68** : 973 ～ 979, 2020］

Corresponding author: *Takao HINOI*, Department of Clinical and Molecular Genetics, Hiroshima University Hospital, Hiroshima 734-8551, Japan. E-mail: thinoi@hiroshima-u.ac.jp

【Key Words】 DNA sequence, RNA sequence, splicing variant, *MUTYH*

Abbreviations: DNA-seq, DNA sequence; RNA-seq, RNA sequence; NGS, next-generation sequencing; CRC, colorectal cancer; MSI, microsatellite instability; LS, Lynch syndrome; dbSNP, Single Nucleotide Polymorphism Database; jMorp, Japanese Multi Omics Reference Panel; InSiGHT, International Society for Gastrointestinal Hereditary Tumors; HGVD, Human Genetic Variation Database

Received November 15, 2019; accepted November 6, 2020

*[1,8]Department of Clinical Laboratory, *[1,7,8]Department of Diagnostic Pathology, *[8]Institute for Clinical Research,
*[9]Honorary President, NHO Kure Medical Center & Chugoku Cancer Center, 3-1 Aoyama-cho, Kure 737-0023, Japan
*[2]Department of Clinical and Molecular Genetics, Hiroshima University Hospital, 1-2-3 Kasumi, Minami-ku,, Hiroshima 734-8551, Japan
*[3 ～ 6,10]Department of Molecular Diagnosis and Cancer Prevention, Saitama Cancer Center, 780 Oazakomuro, Inamachi, Kitaadachi-gun, Saitama-pref. 362-0806, Japan
*[1]Current Affiliation: Department of Clinical Genomic Medicine, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan

I. Backgrounds

The next-generation sequencing（NGS） technique is widely utilized in clinical research and diagnostic tool for somatic and hereditary disorders. We introduced dual analysis of DNA-seq and RNA-seq as a routine protocol for the NGS-based wide detection in genetic analyses. In this report, we show an advantageous example of germline analysis of combining DNA-seq and RNA-seq, which identified two independent splicing variants derived from one single nucleotide alteration. This analytic procedure would provide us with an effective methodology for exploring potential pathogenicity of transcript variants related to the hereditary tumor.

II. Methods and Case Presentation

A. Samples

DNA and RNA were extracted from lymphocytes which were propagated from the peripheral blood mononuclear cells（PBMCs）. PBMCs were harvested from a heparinized whole blood specimen by using Ficoll®-Paque PREMIUM（GE Healthcare, currently Cytiva, MA, USA） and were subsequently cultured in KBM502 （KOHJIN BIO, Saitama, Japan） containing IL-2 supplemented with 10 % of fetal bovine serum （Cytiva）. After one-week or minimally 4-hour of conditioned cell culture, the cells were divided into two tubes; one for the DNA extraction, and another for the RNA extraction with or without puromycin（Thermo Fisher Scientific, Waltham, USA） treatment. Puromycin was applied to inhibit the degradation of abnormal transcripts. DNA and RNA were extracted simultaneously using AllPrep DNA/RNA Mini Kit（Qiagen, Hilden, Germany） after being applied to the standard proteinase K digestion. Extracted DNA was purified with the spin column isolation technique by using a QIAamp DNA Mini Kit（Qiagen）. RNA purification was conducted by using the Isogen phenol-chloroform extraction method according to the manufacture's protocol（Nippon Gene, Tokyo, Japan）. RNA was thereafter synthesized into cDNA with both oligo-dT and random primers for RT-PCR following the manufacture's protocol of the SuperScript First-Strand Synthesis System（Thermo Fisher Scientific, Carlsbad, CA, USA）.

B. Microsatellite instability（MSI） testing

Standard MSI analysis was performed by analyzing the tumor and normal DNA from formalin-fixed paraffine embedded（FFPE） samples. MSI status was determined by comparing the PCR products from the tumor and normal samples in sizes and amounts of nine markers including five Bethesda markers: BAT25, BAT26, D2S136, D5S346, D17S250 and an additional four markers: TP53, D2S123, D3S1067, and D18S51. DNA isolation and data analysis were performed using GeneMapper® Software 5（Applied Biosystems） in a ISO 15189-certified and CAP-accredited outsourcing laboratory.

C. Next-generation sequencing（NGS）

Library preparation for NGS was carried out as described previously [1]. NGS analyses were performed by using MiSeq（Illumina, San Diego, USA）. DNA was sequenced using QIAseq Targeted DNA Custom Panels（Qiagen, Hilden, Germany） according to the manufacturer's instruction. For DNA sequencing, the laboratory-developed targeted panel MHCRCv2（Saitama Cancer Center, Saitama, Japan） was applied covering 15 genes including: *MLH1*, *MSH2*, *MSH6*, *PMS2*, *MSH3*, *PMS1*, *EPCAM*, *MLH3*, *APC*, *MUTYH*, *POLD1*, *POSLE*, *TP53*, *AXIN2*, and *BMPR1A*. Transcripts of the target genes including MMR genes were amplified by PCR with the complementary DNA（cDNA） and sequenced using Nextera XT DNA Library Prep Kit（Illumina） according to the manufacturer's instruction. NGS analysis was conducted on whole coding lesions including all exons covering 5 base pairs upstream or downstream of the exon-intron boundaries for the splice-site alteration analysis.

D. Sequence data analysis

The sequence data were processed by using CLC Genomics Workbench（Qiagen）. Nucleotide numbering of *MUTYH* c.934-2A>G was based on the National Center for Biotechnology Information［NCBI］（https://www.ncbi.nlm.nih.gov） reference sequence NM_001128425.1. Single nucleotide variants, insertions, and deletions were annotated through the mapping and the read alignment to a human genome 19 reference（hg19）. The possible germline variants were annotated by referring to the latest ClinVar after excluding common variants of which general population frequencies are over 0.5 % in Genome Aggregation

Database(gnomAD), the Japanese Multi Omics Reference Panel(jMorp), or the Human Genetic Variation Database(HGVD). The coverage rate was calculated by the visual observation. Curation was conducted by the expert panel discussion consisted with the bioinformaticians and the board certified clinical geneticists in our institute by referring to the American College of Medical Genetics and Genomics(ACMG) 2015 guidelines.

E. Case presentation

A Japanese woman in her 40s was referred to our department for the treatment of an advanced ascending colon tumor. Right hemicolectomy with lymph node dissection was performed for the colon cancer at pathological stage II. The tumor was mucinous adenocarcinoma showing aberrant diffuse positive staining for p53 in the nucleus of tumor cells(**Fig. 1A**). Lynch-syndrome(LS) was first suspected due to the histological type, the location of the tumor, and the familial history harboring, at least, two cases of LS-associated malignancy within second-degree relatives(**Fig. 1B**). All these factors met the revised Bethesda guidelines, a selective screening criterion for LS [2]. Microsatellite instability(MSI) testing with the use of a polymerase-chain-reaction(PCR) was performed as an initial advanced screening according to the recommended algorithm of LS testing [3]. By analyzing nine microsatellite markers including five Bethesda markers, no clear evidence of high-frequency MSI(MSI-H) was detected except in one marker, D3S1067 (**Fig. 1C**). Being consistent with MSI testing, immunohistochemical study showed no abnormal deficiency in mismatch repair protein(MMR) expressions, indicating less association with LS(data not shown). We then proceeded with the genetic testing by using the multi-gene panel MHCRCv2 to investigate germline variants that might be related with young-age onset. As the results of our analysis flows as

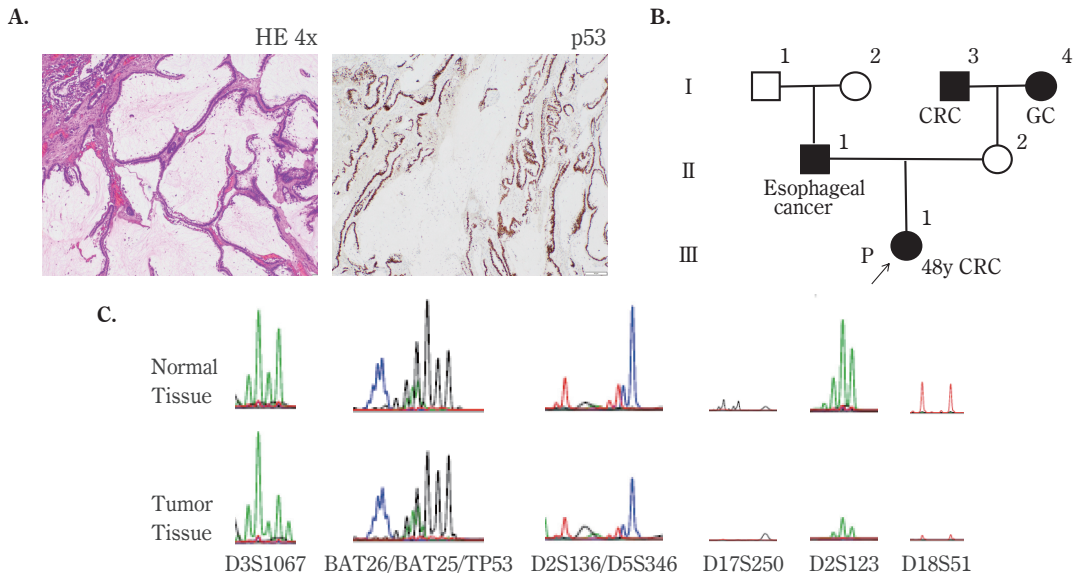

Figure 1 Initial examination strategy of this case suspected for Lynch syndrome.

A: Pathological examination of the resected colon cancer revealed mucinous carcinoma with diffuse positive staining for p53.

B: Family pedigree. There are two diagnosed cases in the second-degree relative; a case of colorectal cancer and a case of gastric cancer, both of which are Lynch syndrome-related tumors. Arrowhead, proband; CRC, colorectal cancer; GC, gastric cancer.

C: Microsatellite instability(MSI) test was performed by comparison with the size of PCR products between the normal and the tumor tissues in nine markers(D3S1067, BAT26, BAT25, TP53, D2S136, D5S346, D17S250, D2S123, and D18S51). A shift was detected in D3S1067. PCR amplification of D17S250 and D18S51 in the tumor tissues was not sufficient for the evaluation.

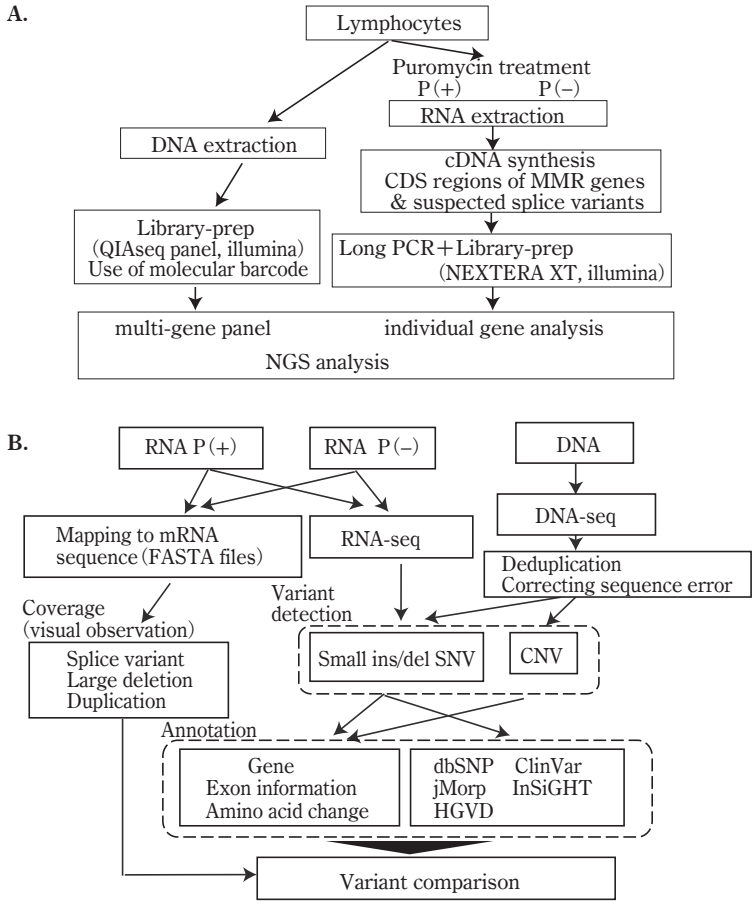

Figure 2 Overview of genetic testing.

A: Testing workflow of NGS analysis in this study. P(+), with puromycin treatment; P(−), without puromycin treatment. cDNA, complementary DNA; CDS, coding region.

B: Workflow of sequence data analysis. RNA P(+), RNA extracted from lymphocytes treated with puromycin; RNA P(−), RNA extracted from lymphocytes without puromycin treatment; ins, insertion; del, deletion; SNV, single nucleotide variant; CNV, copy number variation; dbSNP, ClinVar, jMorp, InSiGHT, and HGVD are representative public-domain reference databases.

shown in **Fig. 2A** and **2B**, two heterozygous single nucleotide germline variants were identified in *TP53* and *MUTYH* through DNA-seq. A variant detected in *TP53* was c.145G>C, leading to a missense of p.Asp49His(D49H)(data not shown), of which germline variant has been reported in the possible relation with pathogenicity [4)5)]. Another variant was a single nucleotide alteration of c.934-2A>G in the exon 10/intron 11 boundary of *MUTYH*, encoding a base excision DNA repair enzyme. RNA analysis was simultaneously performed by cDNA sequencing of reverse-transcription-PCR(RT-PCR) products. It was revealed that a heterozygous *MUTYH* c.934-2A>G resulted in the production of two alternative transcript variants. One variant was a transcript which retained intron 10 harboring a premature stop codon within intron 10 resulting in p.Glu313Serfs*8 due to a stop codon and frame-shifting(**Fig. 3A**, **B**). Another alternative transcript was created by a splice-out of intron 10 with an adjunctive nine-base of the 5'-site in exon 11 as a result of the read-through transcript of intron 10 to the alternative splice acceptor site of exon 11(**Fig. 4A**, **C**). Three-amino-acid shorter form is created by the extra splice-out of nine base pairs in exon 11 ad-

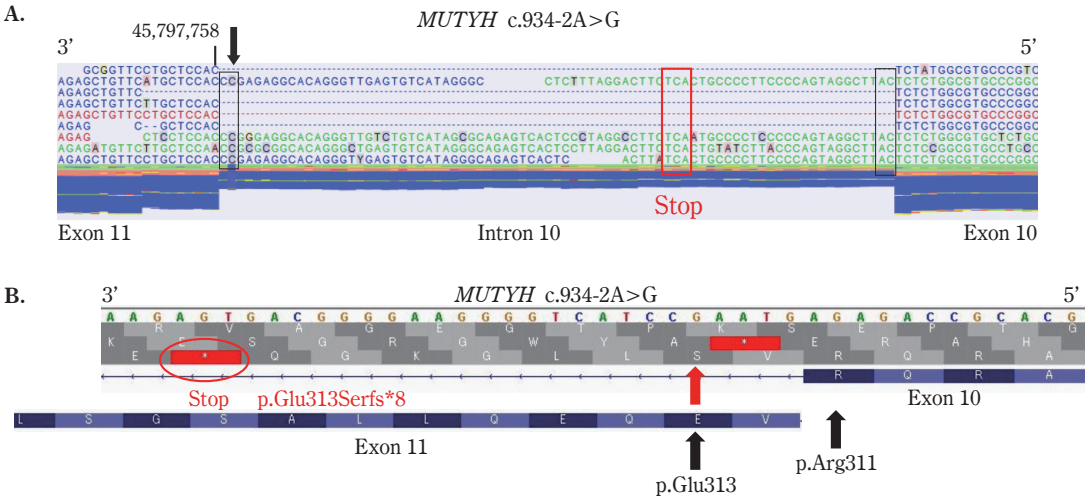

Figure 3 Germline sequence analyses of *MUTYH* RNA.

A: RNA sequence analysis of *MUTYH*. The black arrow indicates the position of one-base alteration of c.934-2A>G in the 3' splice acceptor site in intron 10. The black squares indicate the position of the splice donor site and the splice acceptor site. The position of premature stop codon located in the middle of intron 10 was marked with a red square. **Note:** The nucleotide sequence is complementary.

B: Protein product of *MUTYH* is truncated due to the splicing failure and a stop codon within the intron 10, resulting in a shorter form of p.Glu313Serfs*8; Eight amino acids are translated adjunctively following the translated product of exon 10.

junct to intron 10(**Fig. 4B**). The latter form of variant was less detected in RNA isolated from lymphocytes without puromycin treatment(data not shown), suggesting vulnerability of the additional splice form of the 3'-site of intron 10 of *MUTYH*. The frequency of each type of splicing variants was estimated as equal to 15-25 %(data not shown). These two variants were confirmed as germline heterozygous variants through the combined analysis of DNA and RNA extracted from the lymphocytes.

III. Discussion

We showed a beneficial example of the combined DNA-seq and RNA-seq as a comprehensive variant analysis including the identification of unknown variants that could not be identified by DNA-seq alone. Significance in the dual analysis of DNA and RNA was an identification of two transcriptional forms due to splicing alteration derived from a single nucleotide variant of *MUTYH*. Although it is beneficial procedure with high sensitivity, a point to be considered in the dual analysis as a routine protocol would be time and cost effectiveness. As shown in our previous study [1], RNA-based NGS analysis is reasonably practical as a clinical examination with consistently short duration of analysis as long as 4 days from the nucleic acid extraction. The expense was as low as US$144(as of 2016) per sample in the analysis over 900 specimens [1]. Another consideration would be the types of material selected for germline analysis. The analyses of DNA and RNA extracted from lymphocyte might not necessarily represent the exact expression of two transcripts of *MUTYH* in target organs and cells. As for the genetic implication of the inactivating *MUTYH* germline mutation, either homozygous or compound heterozygous changes are known to be associated with gastrointestinal adenomas and carcinomas of familial polyposis regardless of the lack of the adenomatous polyposis coli gene(*APC*) mutation(OMIM# 608456)[6]. The monoallelic germline *MUTYH* mutant may also increase the risk of specific types of malignancy [7]. The pathogenicity of c.934-2A>G of *MUTYH* is currently a subject of debate in interpretation in ClinVar(Variation ID: 41766). Further investigation and clinical surveillance are necessary to arrive at conclusive determinations of the clinical significance of each transcript variant derived from a c.934-2A>G of *MUTYH*.

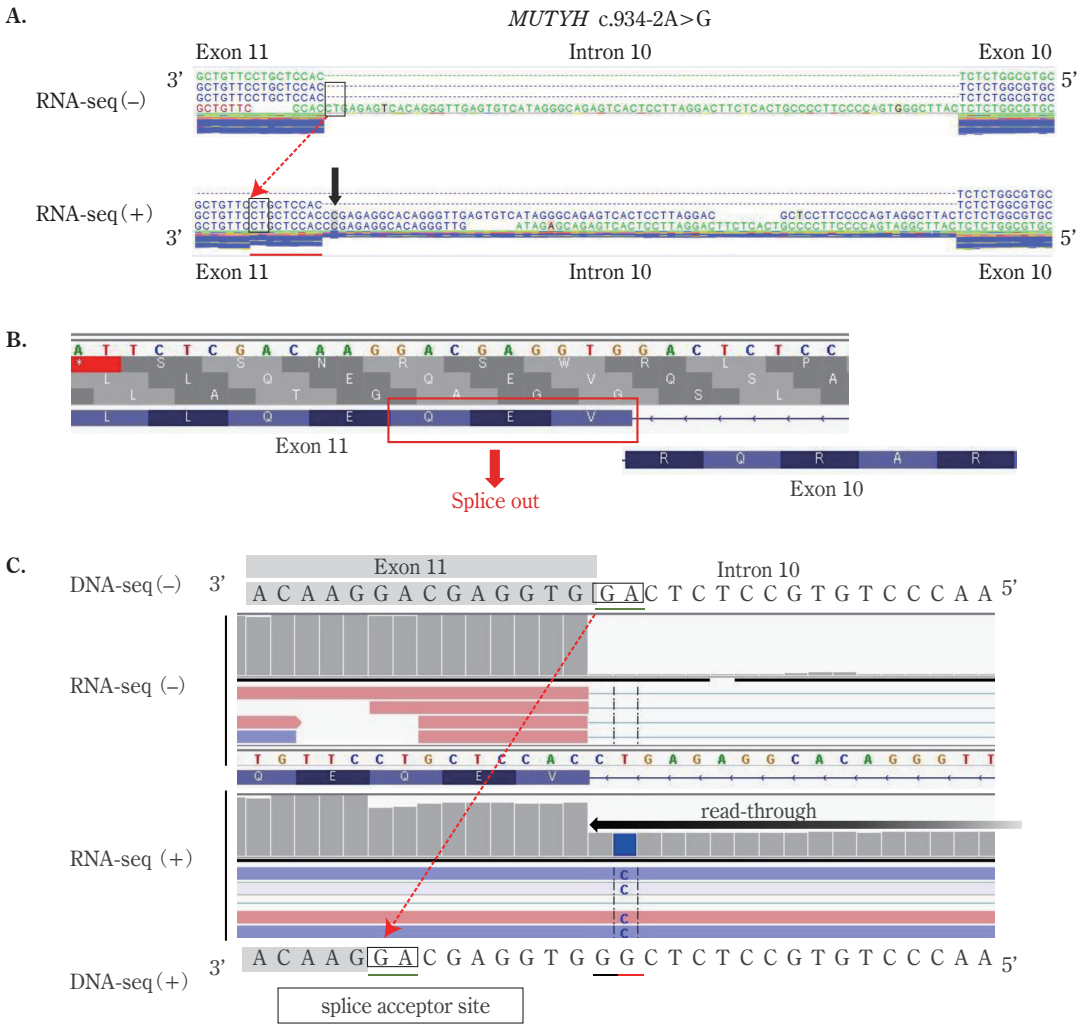

Figure 4 Another alternative splicing variant of *MUTYH*, identified through the analysis of puromycin-treated samples.

A: The sequence analysis for *MUTYH*, by using RNA extracted from lymphocytes without the splice variant (RNA-seq(–), upper panel) or with the splice variant (RNA-seq(+), lower panel). The black arrow indicates the location of a single nucleotide substitution variant of c.934-2A>G. The red under-bar marks the location of the sequence whose depth of coverage was reduced, indicating the possibility of another splicing variant. The black square indicates the location of the functional splice acceptor sites in each sequence.

B: Another protein product of *MUTYH* is due to the additional splice out of three amino acids encoded by the 5'-site of exon 11.

C: Schema of another splicing variant structure, uncovered by the dual analyses of DNA-seq and RNA-seq. DNA-seq(–), DNA sequence analysis of the non-variant form; RNA-seq(–), RNA sequence analysis of the non-variant form; DNA-seq(+), DNA sequence analysis of the variant form; RNA-seq(+), RNA sequence analysis of the variant form.

Acknowledgements

The authors acknowledge the work of past and present members at our departments and technical supports by Keisuke Tanaka and Mari Kikuchi. This research was supported by Japan Agency for Medical Research and Development（AMED） under grant JP 18kk0205004, and JSPS KAKENHI Grant Number JP18K07339. The manuscript proofing was assisted by Naoko Kishida.

Ethics approval and consent to participate

All procedures were followed in accordance with the ethical standards of the responsible committee on human experimentation and with the 1975 Helsinki declaration, as revised in 2013, and the Japanese ethical guidelines for human genome/gene analysis research. This study was approved by the Institutional Review Boards of Saitama Cancer Center（no. 729） and NHO Kure Medical Center（no. 29-36）. Written consent was obtained from the patient before inclusion in the study.

Disclosure

The authors declare that there are no conflicts of interest associated with the present study.

References

1） Yamamoto G, Kikuchi M, Kobayashi S, et al. Routine genetic testing of lung cancer specimens derived from surgery, bronchoscopy and fluid aspiration by next generation sequencing. Int J Oncol 2017; 50（5）: 1579-89.
2） Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer（Lynch syndrome） and microsatellite instability. J Natl Cancer Inst 2004; 96（4）: 261-8.
3） Sinicrope FA. Lynch Syndrome-Associated Colorectal Cancer. N Engl J Med 2018; 379（8）: 764-73.
4） Yamaguchi K, Urakami K, Nagashima T, et al. Prevalence of low-penetrant germline *TP53* D49H mutation in Japanese cancer patients. Biomed Res 2016; 37（4）: 259-64.
5） Yamazaki F, Shima H, Osumi T, et al. Nodular lymphocyte-predominant Hodgkin lymphoma in a 15-year-old boy with Li-Fraumeni syndrome having a germline *TP53* D49H mutation. J Pediatr Hematol Oncol 2018; 40（3）: e195-7.
6） Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of *MYH* associated with somatic G:C → T:A mutations in colorectal tumors. Nat Genet 2002; 30（2）: 227-32.
7） Kline CN, Joseph NM, Grenert JP, et al. Inactivating *MUTYH* germline mutations in pediatric patients with high-grade midline gliomas. Neuro Oncol 2016; 18（5）: 752-3.