Evaluation of an amplicon-based custom gene panel for the diagnosis of hereditary tumors

S. SHINRIKI[1,2,*], M. MAESHIRO[2,3], K. SHIMAMURA[1,2], J. KAWASHIMA[4], E. ARAKI[4], M. IBUSUKI[5], Y. YAMAMOTO[5], H. IWASE[5], Y. MIYAMOTO[6], H. BABA[6], M. YAMAGUCHI[7], H. MATSUI[1,2,*]

[1]Department of Molecular Laboratory Medicine, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [2]Department of Laboratory Medicine, Kumamoto University Hospital, Kumamoto, Japan; [3]Department of Oral and Maxillofacial Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [4]Department of Metabolic Medicine, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [5]Department of Breast and Endocrine Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [6]Department of Gastroenterological Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [7]Amelieff Corporation, Tokyo, Japan

*Correspondence: hmatsui@kumamoto-u.ac.jp; satoru.shinriki@gmail.com

Genetic testing based on next-generation sequencing (NGS) analysis has recently been used to diagnose hereditary diseases. In this study, we explored the usefulness of our custom amplicon panel that targeted 23 genes related to hereditary tumors given in the American College of Medical Genetics and Genomics recommendations. We applied our custom NGS panel to samples from 12 patients previously diagnosed by Sanger sequencing as having the diseases or diagnosed clinically by meeting the diagnostic criteria in this study. Our gene panel not only successfully identified all variants detected by Sanger sequencing but also identified previously unrecognized variants that resulted in confirmation of the disease, or even in the revision of the diagnosis. For instance, a patient identified with an *SDHD* gene mutation actually had von Hippel-Lindau (VHL) syndrome, as determined by the presence of a pathogenic *VHL* gene variant. We also identified false-positive results that were generated by amplification of genome regions that are not intended to be investigated. In conclusion, NGS-based amplicon sequencing is a highly effective method to detect germline variants, as long as they are also carefully reviewed by manual inspection.

Key words: hereditary tumors, gene panel testing, amplicon sequencing, next-generation sequencer, germline variants

With the improvement of genome data acquisition throughput and the reduced cost of such analysis, gene panel testing via massive parallel sequencer systems (next-generation sequencer, NGS) is becoming widely used to diagnose various diseases caused by genome abnormality. Also, gene mutation data obtained via NGS were recently used in companion diagnostics, in which the indication of a molecular targeted drug is determined by the presence of gene mutations. For instance, the Oncomine DX Target Test [1], which analyzes 23 genes related to non-small cell lung cancer, was approved as a companion diagnostic by the U.S. Food and Drug Administration (FDA) in 2017 and by the Pharmaceuticals and Medical Devices Agency in Japan in 2018. The Foundation One CDx test for cancer genome profiling was also approved by the FDA in 2017 [2]. NGS has a big advantage in that it can analyze a large genome region at once, so it is likely to be especially useful in isolating candidate gene alterations for diseases in which multiple respon-

sible genes are supposedly involved. NGS is also useful to identify the loss-of-function type of mutations in which the position at which a mutation occurs varies among patients, especially for large genes.

In this study, we evaluated the usefulness of our custom gene panel for 23 genes associated with hereditary tumors. To this end, we analyzed 12 peripheral blood specimens obtained from patients who were previously diagnosed with hereditary tumors via genetic testing that utilized Sanger sequencing or patients who were determined clinically as having the diseases, i.e. those who met the diagnostic criteria of the diseases. For example, as the Results section shows, a patient who was previously diagnosed to have hereditary pheochromocytoma-paraganglioma syndrome (HPPS) because of detection of a succinate dehydrogenase (SDH) mutation [3, 4] actually had a pathogenic germline *VHL* mutation according to our gene panel analysis, which led to a corrected diagnosis of von Hippel-Lindau (VHL) syndrome

[5, 6]. As this example illustrates, when more than two separate syndromes that may manifest overlapping phenotypes occur, a specific disease may sometimes be difficult to diagnose by using conventional Sanger sequencing in terms of time and cost of analyzing all responsible candidate genes. NGS-based analysis may be able to resolve this issue.

In the current traditional NGS system in which short-read sequence tags of at most 300 base pairs long are aligned to the reference genome, certain points about this method should be noted. In other words, simple reporting of candidate gene alterations produced from the sequencer should be avoided. Potential mutations must be manually reviewed by using a genome browser to confirm that a mutation is a true mutation as opposed to a false-positive result. For example, the *PMS2* gene, which we included in our gene panel, has pseudogenes, including *PMS2CL* and *PMS2P3*, and the presence of these pseudogenes has been shown to impede the analysis of *PMS2* gene mutation. A long-range polymerase chain reaction (PCR) or other equivalent methods that can distinguish between *PMS2* and its pseudogenes has been widely used to search for *PMS2* gene mutations [7–9]. However, when the short-read amplicon sequencing method is applied to *PMS2* analysis, probes for *PMS2* may mis-anneal to and amplify the pseudogenes or such mis-amplified amplicons are misaligned to *PMS2*, which gives rise to an incorrect interpretation of the gene mutation. We show such data in Results and discuss later.

In this manuscript, we reveal the usefulness of gene panel testing for hereditary tumors by means of NGS. In addition, we discuss the importance of considering the minor allele frequency (MAF) and the characteristics of each amplicon so that we can carefully evaluate the authenticity and clinical significance of candidate gene mutations.

Patients and methods

Patients. For this study, we obtained samples from 12 patients, previously diagnosed of having hereditary tumors, and a sample from a healthy volunteer (38 year old man). Diagnoses were made according to the diagnostic criteria for each disease, primarily via a genetic method – Sanger sequencing – and according to phenotypic criteria, at medical departments in Kumamoto University Hospital, or at hospitals that referred the study patients to Kumamoto University Hospital. Table 1 summarizes the characteristics of the patients.

After we obtained written informed consent from the patients, we collected peripheral blood samples, and we then extracted genomic DNA from lymphocytes by using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The quantity of the DNA was measured by Bioanalyzer (Agilent, Santa Clara, CA), and the quality was evaluated as DNA integrity number (DIN) value. The DNA was then subjected to sequencing library preparation.

The use of patients' DNA for this purpose was approved by the Ethics Committee for Human Genome Research of Kumamoto University (Approval number: 354), and the study was performed in accordance with the Declaration of Helsinki.

Panel design. To characterize germline gene mutations associated with hereditary tumors, we designed a custom gene panel that targeted 23 genes. The genes were selected according to the American College of Medical Genetics and Genomics (ACMG) recommendations for reporting of incidental findings but excluded genes not related to hereditary tumors [10]. The amplicons for the gene panel were designed by Design Studio (Illumina, San Diego, CA). Table 2 lists the target genes. All coding exons of the 23 genes were designed to be amplified with the following parameters: reference genome: GRCh37/hg19, variant source: 1000 Genomes, amplicon length: 175 bp, exon padding: 25 bp. A total of 827 amplicons with cumulative targets of 73,112 bp (with a total gap distance of 614 bp) were designed.

Library preparation and parallel sequencing. The sequencing library was prepared by using a TruSeq Custom Amplicon Low Input (TSCA-LI) Kit (Illumina) according to the manufacturer's instructions. Briefly, 10 ng of non-fragmented genomic DNA was first hybridized with the custom oligo probes that contained upstream and downstream oligo-DNA probes complementary to the designed regions. The hybridized upstream probes were then complementarily extended by a DNA polymerase through targeted regions, which was followed by ligation to the 3' end of the downstream probes. Finally, synthesized DNA with universal sequences at both ends was amplified via PCR with primers with an index and a sequence required for cluster formation. The amplified products were measured with a Bioanalyzer (Agilent) to determine whether they were within the expected sizes and concentrations, after which they were stored as sequencing libraries.

The prepared libraries mixed in equal amounts were subjected to parallel sequencing on a MiSeq platform (Illumina) by using the MiSeq Reagent kit v2 for 151 cycles at paired-end mode. The PhiX control library (Illumina) was spiked-in during each run at a final concentration of 12 pM to estimate the sequencing error rate, as described in the manufacturer's protocol.

Bioinformatics. MiSeq reporter v2.6.2 was used for data analysis, including sequence alignment to the reference genome (GRCh37/hg19), single-nucleotide variant calling, and insertion/deletion (indel) detection. As Figure 1 shows, each variant call format (VCF) file generated by the MiSeq reporter was then subjected to Base Space Variant Interpreter (Illumina). The annotated variants were first filtered against a depth of <30 at each position, intronic and synonymous variants, and low variant allele frequency. Splice site variants were retained. The remaining variants were subsequently filtered for an MAF of >0.02 in the population-based databases including the Genome Aggregation Database (gnomAD), the

Exome Aggregation Consortium (ExAC), 1000 Genomes, and human genetic variation database (HGVD) [11], which is a Japanese variant database. In this study, the MAF filter was not applied to the variants detected in previous genetic tests so that we could compare the previous variants with candidate variants obtained by our NGS analysis (Table 2). The remaining variants were manually and carefully reviewed by using an Integrative Genomics Viewer (IGV) [12, 13]. Variants also detected in a sample that was obtained from the healthy volunteer were removed. For interpretation of the variants, we used the following public databases (last access: March 2019): ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) and COSMIC (https://cancer.sanger.ac.uk/cosmic). Then, we conducted a literature review of the clinical significance of the remaining variants of uncertain significance (VUS) by using PubMed and public databases, which was followed by discussions with attending physicians about the validity of a variant as the cause of a patient's disease. On the basis of the patients' detailed clinical manifestations, we performed a literature search again. Sanger sequencing was used to verify the variants as necessary. Finally, we report on variants including VUS with detailed information about the bioinformatics analysis and clinical significances.

Results

Participant characteristics. Table 1 provides detailed characteristics of the 12 patients included in this study.

According to clinical criteria, they had been diagnosed as having hereditary tumor syndromes including multiple endocrine neoplasia (MEN) type 1 (n=5), MEN type 2 (n=2), hereditary breast and ovarian cancer (HBOC, n=1), HBOC carrier (n=1), familial adenomatous polyposis (FAP, n=1), HPPS/VHL (n=1), and Lynch syndrome (n=1). Of six patients (UPN1, 6, 8–10, and 12) whose family histories were not typical, five patients other than UPN12 had relatives with benign or malignant tumors. All cases except for the FAP case (UPN5) had been genetically tested by DNA Sanger sequencing. For one breast cancer patient (UPN2), no pathogenic *BRCA* mutation had been reported, and this patient was tentatively diagnosed as having Lynch syndrome. We analyzed the genomic DNA in peripheral blood samples from these patients and a healthy volunteer via NGS with our custom hereditary tumor panel and analysis pipeline.

Sequence quality control and coverage. We developed a custom hereditary tumor NGS panel covering all coding exons of 23 genes associated with hereditary tumor syndromes (Table 2; see also Patients and methods). Figure 1 shows our analytical workflow system that utilizes this panel consisting of sample preparation, sequencing, and data analysis.

First, we monitored sequencing run quality by using parameters including cluster density, cluster passing filter, Q30, and error rate based on the PhiX spike-in. Every run was performed with reasonable values for all parameters: average cluster density (K/mm^2) between 620 and 990, cluster
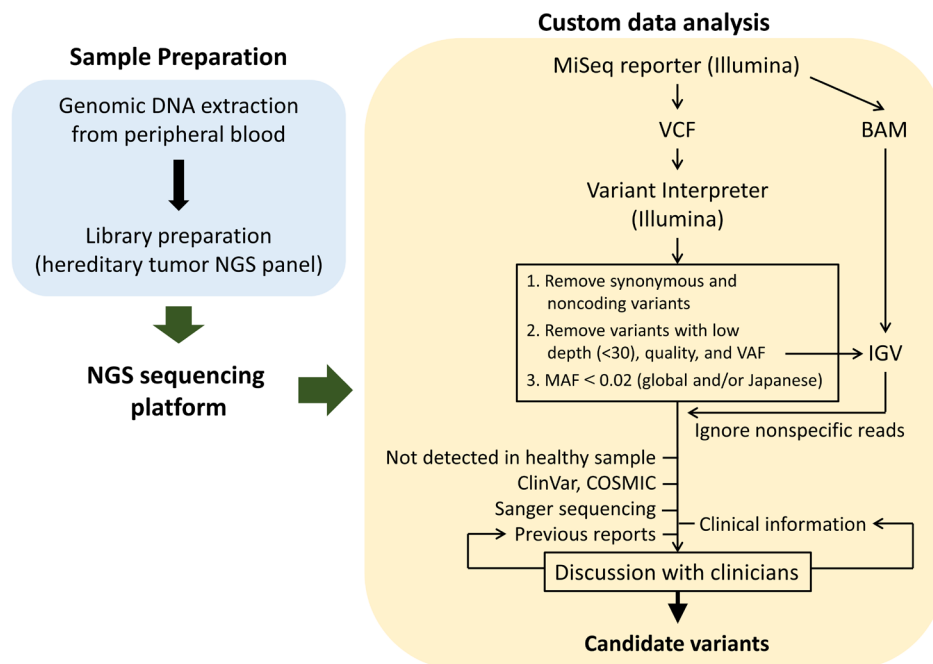
Figure 1. Diagnostic workflow with our hereditary tumor NGS panel. The workflow consists of sample preparation, sequencing, and our custom data analysis process. The TSCA-LI system was used for library preparation, and sequencing was performed with the MiSeq instrument. In the data analysis process, VCF and BAM files generated by the MiSeq reporter were carefully analyzed. The MAF filter was not applied to variants detected in previous genetic tests. VAF – variant allele frequency

Table 1. Participant characteristics and variants detected.

Patient# (UPN)	Sex	Tumor history (proband and 1st- to 3rd-degree relatives)	Clinical criteria	Previous analysis	Coding infrequent non-synonymous variants as detected by hereditary tumor NGS panel analysis	MAF	ClinVar*
1	F	Proband: breast cancer (left, 36y; right, 45y), hyperparathyroidism (19y), primary aldosteronism (35y). Mother and maternal grandmother: endometrial cancer. Maternal great aunt 1: breast cancer. Maternal great aunt 2: cervical cancer. Maternal great aunt 3-5: cancer (origins unknown).	HBOC	BRCA1 (NM_007300.3): c.3607C>T, p.Arg1203Ter, hetero	BRCA1 (NM_007300.3): c.3607C>T, p.Arg1203Ter, hetero	0.00001 (gnomAD)	Pathogenic
2	F	Proband: breast cancer (35y). Mother: ovarian cancer. Maternal aunt 1: colon cancer. Maternal aunt 2: pancreatic cancer. Maternal uncle 1 and 2: lung cancer. Maternal cousin: breast cancer. Paternal aunt: colon cancer.	Lynch syndrome (suspected)	BRCA1/2 genetic analysis negative	None	–	–
3	F	Proband: pheochromocytoma (left, 35y; right, 51y), thyroid medullary carcinoma (52y). Sister and 3rd son: pheochromocytoma, thyroid medullary carcinoma. 2nd son: thyroid medullary carcinoma.	MEN2	RET (NM_020975.4): c.1901G>A, p.Cys634Tyr, hetero	RET (NM_020975.4): c.1901G>A, p.Cys634Tyr, hetero	No data	Pathogenic
					RET (NM_020975.4): c.833C>A, p.Thr278Asn, homo	0.00211 (gnomAD), 0.0200 (HGVD)	Conflicting interpretations of pathogenicity
4	F	Proband: pheochromocytoma (right, 23y; left, 32y), pancreatic endocrine tumor (32y), spinal cord hemangioblastoma (suspected). Mother: duodenal carcinoid. 2nd son: pheochromocytoma.	HPPS/ VHL	RET: negative; SDHD (NM_003002.3): c.331G>A, p.Val111Ile, hetero	SDHD (NM_003002.3): c.331G>A, p.Val111Ile, hetero	0.00001 (gnomAD), 0.0036(HGVD)	Uncertain significance
					VHL (NM_000551.3): c.496G>T, p.Val166Phe, hetero	No data	Pathogenic
					WT1 (NM_024426.4): c.1249+7G>C, hetero	No data	No data
5	F	Proband: colon polyps (<100; 20s y) and colon cancer. Father, paternal uncle, and paternal cousin: colon cancer.	FAP	None	APC (NM_000038.5): c.1172_1179delTTCACTCA, p.Ile391ThrfsTer3	No data	No data
					APC (NM_000038.5): c.3378C>G, p.Ser1126Arg, hetero	0.00004 (gnomAD)	Conflicting interpretations of pathogenicity
6	F	Proband: primary hyperparathyroidism (42y), nodular goiter. Mother: brain tumor.	MEN1	MEN1 (NM_130803.2): c.974C>T, p.Pro325Leu, hetero	MEN1 (NM_130803.2): c.974C>T, p.Pro325Leu, hetero	No data	Pathogenic
					MLH1 (NM_000249.3): c.1990-6G>A, hetero	0.00006 (gnomAD), 0.0104(HGVD)	Likely benign
7		Proband: primary hyperparathyroidism (61y), pancreatic endocrine tumor, and adrenal gland tumor (62y). 1st son: primary hyperparathyroidism, pituitary tumor. 2nd son: primary hyperparathyroidism.	MEN1	MEN1 (NM_130803.2): c.793C>T, p.Gln265Ter, hetero	MEN1 (NM_130803.2): c.793C>T, p.Gln265Ter, hetero	No data	Pathogenic
8		Proband: thyroid medullary carcinoma (50y). Mother: pancreatic cancer, liver cancer.	MEN2	RET (NM_020975.4): c.2071G>A, p.Gly691Ser, hetero	TSC2 (NM_000548.3): c.856A>G, p.Met286Val, hetero	0.00180 (gnomAD), 0.0156 (HGVD)	Likely benign

Table 1. *Continued.*

Patient# (UPN)	Sex	Tumor history (proband and 1st- to 3rd-degree relatives)	Clinical criteria	Previous analysis	Coding infrequent non-synonymous variants as detected by hereditary tumor NGS panel analysis	MAF	ClinVar*
9	F	Proband: none (24y). Maternal aunt and grandmother: ovarian cancer.	HBOC carrier	BRCA1 (NM_007300.3): c.2800C>T, p.Gln934Ter; hetero	BRCA1 (NM_007300.3): c.2800C>T, p.Gln934Ter, hetero	No data	Pathogenic
					BRCA2 (NM_000059.3): c.8165C>T, p.Thr2722Ile, hetero	No data	Uncertain significance
					MUTYH (NM_001128425.1): c.74G>A, p.Gly25Asp, hetero	0.01051 (gnomAD), 0.0054 (HGVD)	Conflicting interpretations of pathogenicity
					MUTYH (NM_001128425.1): c.53C>T, p.Pro18Leu, hetero	0.00104 (gnomAD), 0.0054 (HGVD)	Conflicting interpretations of pathogenicity
10	M	Proband: gastrinoma, acromegaly (51y). Father: thymoma.	MEN1	MEN1 (NM_130803.2): c.1636G>A, p.Ala546Thr, hetero	None		
11	F	Proband: pancreatic endocrine tumor (19y). Father: malignant gastrinoma. Sister: primary hyperparathyroidism, pancreatic endocrine tumor.	MEN1	MEN1 (NM_130803.2): c.927+2T>G, hetero	MEN1 (NM_130803.2): c.927+2T>G, hetero	No data	No data
					PMS2 (NM_000535.5): c.379G>A, p.Ala127Thr, hetero	0.00046 (gnomAD), 0.0056 (HGVD)	Benign/likely benign
12	M	Proband: insulinoma (18y), pancreatic endocrine tumor, primary hyperparathyroidism, pituitary tumor (35y).	MEN1	MEN1 (NM_130803.2): c.1025C>A, p.Ala342Asp, hetero	MEN1 (NM_130803.2): c.1025C>A, p.Ala342Asp, hetero	No data	No data
					RET (NM_020975.4): c.874G>A, p.Val292Met, hetero	0.00057 (gnomAD), 0.0015 (HGVD)	Conflicting interpretations of pathogenicity
					RET (NM_020975.4): c.200G>A, p.Arg67His, hetero	0.00102 (gnomAD), 0.0041 (HGVD)	Likely benign
					RET (NM_020975.4): c.2944C>T, p.Arg982Cys, hetero	0.01927 (gnomAD), 0.0048 (HGVD)	Benign
					NTRK1 (NM_002529.3): c.892C>T, p.His298Tyr, hetero	No data	No data
					PMS2 (NM_000535.5): c.1688G>T, p.Arg563Leu, hetero	0.00623 (gnomAD), 0.0004 (HGVD)	Benign

*Data were collected from ClinVar in March 2019. Abbreviations: hetero - heterozygous; homo - homozygous.

passing filter of >91.7%, Q30 of >86%, and error rate of <0.45%. These results indicated that every sequencing run performed was of sufficient quality.

To analyze the sequencing quality at each genome position, we utilized the coverage per amplicon and the mean coverage per sample as proxies. These coverage values were automatically generated for every sample in the run via a coverage file that we developed in-house (Supplementary Table S1). The sequencing runs produced an average of 2.0×10^6 (0.73–8.2×10^6) paired reads. Coverage breadth was specified as the proportion of unique reads that were mapped unambiguously to target regions. Our evaluation of coverage breadth revealed that amplicons with the depth of ≥30, which is a recommended minimal threshold to provide reliable diagnostic analysis of germline mutations in blood-derived samples [14], occupied 91.9% (90.5–100%) of all amplicons (Figure 2A). In addition, the coverage uniformity was 90.8±0.4% (Supplementary Table S1). Therefore, the majority of exons were sequenced at a reasonable depth for germline variant calling. Variants with a depth of <30 were regarded as non-informative, and we defined amplicons with a depth of <30 in more than 75% of the samples as repeatedly failing ones. We found 8 repeatedly failing regions in *TSC2* (4/75 amplicons: 5.3%, Supplementary Figure S1), *MSH6* (2/48: 4.2%), *STK11* (1/15: 6.7%), and *WT1* (1/19: 5.3%) genes (Supplementary Table S1). Next, we assessed the relationship between guanine-cytosine (GC) content and amplicon coverage. We found no apparent correlation of GC content with mean normalized coverage depth (Figure 2B). Although all the repeatedly failing regions had relatively high GC

Table 2. Target genes in our hereditary tumor NGS panel.

Genes	Syndromes	Associated tumors
BRCA1	HBOC	Breast cancer, ovarian cancer
BRCA2	HBOC	Breast cancer, ovarian cancer
TP53	Li-Fraumeni syndrome	Osteosarcoma, brain tumor, soft tissue sarcoma, and others
STK11	Peutz-Jeghers syndrome	Small intestinal polyps, adenomatous polyposis, and others
MLH1	Lynch syndrome	Colorectal cancer, endometrial carcinoma, ovarian cancer, gastric cancer, skin cancer, and others
MLH2	Lynch syndrome	Colorectal cancer, endometrial carcinoma, ovarian cancer, gastric cancer, skin cancer, and others
MLH6	Lynch syndrome	Colorectal cancer, endometrial carcinoma, ovarian cancer, gastric cancer, skin cancer, and others
PMS2	Lynch syndrome	Colorectal cancer, endometrial carcinoma, ovarian cancer, gastric cancer, skin cancer, and others
APC	FAP	Colorectal cancer and others
MUTYH	MAP	Colorectal cancer and others
VHL	VHL	Brain hemangioma, retinal hemangioma, renal cell carcinoma, pancreatic tumors, and others
MEN1	MEN1	Hyperparathyroidism, pituitary adenoma, neuroendocrine tumor, and others
RET	MEN2	Medullary thyroid carcinoma, pheochromocytoma, hyperparathyroidism, and others
NTRK1	FMTC	Thyroid carcinoma and others
PTEN	PTHS	Thyroid carcinoma endometrial carcinoma, breast cancer, kidney cancer, and others
RB1	Retinoblastoma	Retinoblastoma
SDHD	HPPS	Pheochromocytoma, paraganglioma
SDHAF2	HPPS	Pheochromocytoma, paraganglioma
SDHC	HPPS	Pheochromocytoma, paraganglioma
SDHB	HPPS	Pheochromocytoma, paraganglioma
TSC1	TSC	Angiofibroma, brain tumor, kidney tumor, and others
TSC2	TSC	Angiofibroma, brain tumor, kidney tumor, and others
WT1	Wilms' tumor	Wilms' tumor

MAP – MUTYH-associated polyposis; FMTC – familial medullary thyroid carcinoma; PTHS – PTEN hamartoma tumor syndrome; HPPS – hereditary pheochromocytoma-paraganglioma syndrome; TSC – tuberous sclerosis complex.
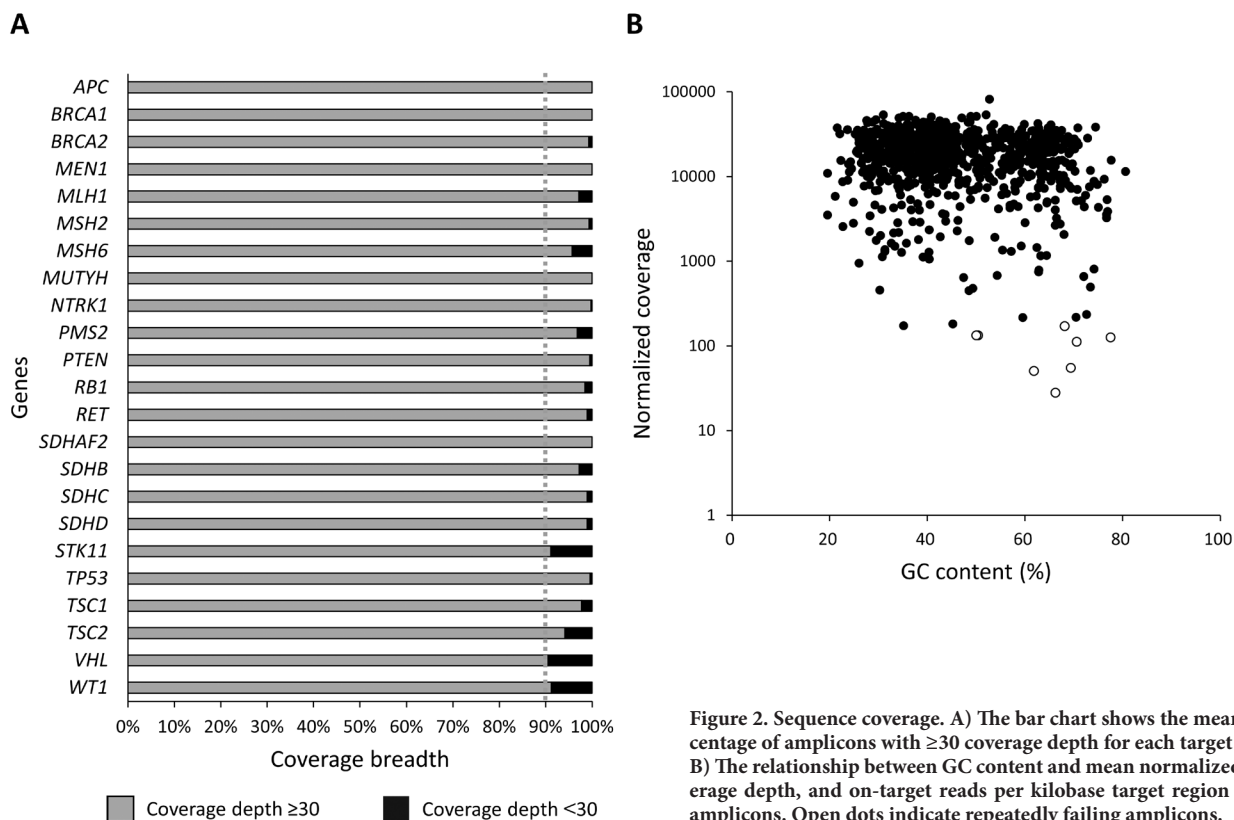

Figure 2. Sequence coverage. A) The bar chart shows the mean percentage of amplicons with ≥30 coverage depth for each target gene. B) The relationship between GC content and mean normalized coverage depth, and on-target reads per kilobase target region of all amplicons. Open dots indicate repeatedly failing amplicons.

contents, this feature was not believed to be the only cause of low coverage depth. Similar results were obtained when coverage depth was normalized by total or mean reads for each sample (data not shown).

Next, to review the structure of the sequenced amplicons, we analyzed BAM files by using IGV, a genome browser, against the remaining variants. Because of our analytical pipeline in which sequence tags were aligned to the whole genome, we found splice site variants in *PMS2CL* (n.947-6C>G) in UPN1, 8, and 10, and *PMS2P3* (n.996+4A>T) in UPN11, both of which were not targeted in our gene panel (Supplementary Figure S2). Our manual inspections via IGV and BLAST showed that the sequences of the corresponding amplicons in *PMS2CL* and *PMS2P3* were identical to the intron 11 to exon 12 and intron 2 to exon 2 in the *PMS2* gene, respectively. Inasmuch as we detected no variant in the corresponding position in *PMS2*, we concluded that the probes mis-annealed to and amplified the *PMS2CL* and *PMS2P3*

regions. As mentioned in the Introduction, *PMS2* has a large number of highly homologous sequences. The *PMS2CL* pseudogene, in particular, has more than 98% sequence identity with the 30 regions of *PMS2*, including exons 9 and 11–15 [15, 16]. Indeed, we found that the amplicons with a depth of ≥30 may be generated by mis-amplification in part of exons 2–5 of *PMS2CL* and in part of exons 5, 7, and 8 of *PMS2P3* (Supplementary Tables S2 and S3). Although our panel did not miss *PMS2* gene mutations, because all coding exons of this gene were amplified and sequenced with enough coverage in all samples (Supplementary Table S1), the specificity for *PMS2* needs improvement.

Variants detected and clinical implications. The average number and standard deviation of initially called variants in VCF files for all samples including the healthy volunteer sample was 154±9. The variants were decreased to 22±3 just before being filtered by MAF (see the filtering process in Figure 1). The non-synonymous infrequent variants detected

Figure 3. The landscape of candidate variants identified by our custom hereditary tumor panel. Non-synonymous infrequent variants detected by NGS analysis with our hereditary tumor panel are shown. The data correspond to the data in Table 2. Style of each cell of the matrix varies according to the type of variant found: plain for missense, horizontal hatching for splice region, and vertical hatching for both missense and frameshift. The number in each cell indicates the number of variants. Bars on the right side represent the total number of variants per gene. The bottom bars represent the total number of variants per sample. The bold solid or dashed frames indicate pathogenic mutations; the dashed frame indicates that one of the two variants in the *RET* gene was identified as pathogenic.

in our panel analysis as possible causal mutations in each patient and detailed information about these variants are given in Table 1 and Supplementary Table S4, which also include all previously identified variants to show why they were removed from candidate status.

NGS analysis with our custom panel correctly detected all 10 previously recognized disease-causing variants, providing a sensitivity of 100% (Table 1 and Supplementary Table S4). However, only half of the variants (in UPN1, 3, 6, 7, and 9) could be classified as pathogenic mutations, and hence the initial diagnosis was not changed for these patients. The remaining half of the variants (in UPN8, 10, 11 and 12) were determined to be VUS, except for that in UPN4 described later. The *RET* p.Gly691Ser previously detected in UPN8 with medullary thyroid carcinoma (MTC) had a high MAF, and the clinical significance of this mutation is uncertain but is considered low [17]. In addition, this variant was detected in UPN12 without a thyroid tumor (data not shown). To our knowledge, the biological relationship between the mutation in *TSC2* (UPN8) and the occurrence of thyroid cancer, and between the mutation in *PMS2* (UPN11) and pancreatic neuroendocrine tumor, has rarely been reported. Virtually no candidate pathogenic variant was observed for UPN2 and UPN10, in which *MEN1* p.Ala546Thr that was previously believed to be pathogenic in UPN10 had a high MAF and indeed was also detected in other 3 patients analyzed in this study. Figure 3 shows a summary of candidate variants identified by our custom hereditary tumor panel.

One of the important findings was that our NGS analyses identified a *VHL* gene mutation (p.Val166Phe), which is known to be a causal mutation of VHL [18–20], in UPN4. This female patient suffered from bilateral pheochromocytoma and a pancreatic endocrine tumor. When the unilateral pheochromocytoma firstly developed, the *RET* gene mutation had been ruled out by Sanger sequencing. The physician in charge screened mutations in *SDHB* and *SDHD* genes when the patient developed bilateral pheochromocytomas and a pancreatic tumor and identified the *SDHD* p.Val111Ile mutation (Table 1). The MAF of this *SDHD* mutation is indeed low, but the clinical significance is still uncertain. Subsequently, her son also developed pheochromocytoma, i.e. inherited pheochromocytoma. An important feature of *SDHD*-related disease is that it is characterized by maternal genomic imprinting [21]. Maternal imprinting is caused by DNA or histone methylation and hence silencing of the maternal allele [22, 23]. Therefore, individuals who inherit a mutation from the mother should not develop such tumors [3, 24]. Given the presence of inherited pheochromocytoma, pancreatic endocrine tumor, and then suspected hemangioblastoma, we concluded that the disease of UPN4 is sporadic VHL and, according to the necessary ethical procedure, we changed the diagnosis to VHL.

UPN5 was diagnosed on a clinical basis (without genetic tests) as having FAP by meeting the diagnostic criteria (multiple adenomas in the colon with a clear family history). Our NGS analysis detected the deletion mutation p.Ile391ThrfsTer3 (c.1172_1179delTTCACTCA) and the missense mutation p.Ser1126Arg in the *APC* gene. The use of Sanger sequencing confirmed the presence of these mutations, but whether these two variants were on the same allele or not, was not confirmed. The deletion that causes the frameshift is located at the distal 3' end of exon 16 of the *APC* gene and is characterized by a deletion of eight nucleotides, which generates a premature stop codon at codon 393 (Supplementary Figure S3). Although we did not analyze the presence of the additional somatic mutation ("second hit") in the tumor lesion or find detailed epidemiological information, we believe this frameshift mutation to be the cause of FAP. Given that this patient had less than 100 polyps, the disease was classified as attenuated FAP (AFAP), which is characterized by a reduced number of polyps (10–100), later age at disease onset, right-side directed distribution of polyps, and lower colorectal cancer risk (up to 70%) compared with classical FAP [25, 26]. Consistent with our finding, mutations associated with AFAP have been found at the 5' end, exon 9, and the distal 3' end of the *APC* gene [27].

UPN12 had typical manifestations of MEN type 1 including a pancreatic endocrine tumor, primary hyperparathyroidism, and a pituitary tumor. Although we detected *MEN1* p.Ala342Asp, also identified in previous genetic tests, this variant remained as a VUS. However, this patient had 4 *RET* gene variants, among which p.Gly691Ser (data not shown), also detected in UPN8 (Table 1 and Supplementary Table S4), has conflicting significance. As an interesting finding, we found a previous report in the literature showing a familial MTC case harboring *RET* p.Val292Met, p.Arg67His, and p.Arg982Cys, the variants that UPN12 also had [28]. Although UPN12 did not show any indication of thyroid tumor, a follow-up with a focus on the thyroid may be needed.

Although the analyses via our hereditary tumor NGS panel presented certain difficulties that would be common to amplicon sequencing methodology and a need for design improvement, we did identify all previously identified variants as well as certain clinically valid new variants, which suggests the usefulness of this panel and our analytical pipeline for the diagnosis of hereditary tumors.

Discussion

Genomic analysis using an NGS platform is an innovative technology that allows a large amount of data to be obtained and analyzed at once, and it has therefore been introduced in many fields from basic research to clinical applications. In clinical use, NGS is being used to diagnose diseases caused by genomic abnormalities as well as to assist in deciding about the use of molecular target drugs; some genetic testing panels have thus been approved for clinical use [1, 2]. However, as many experts in this field recognize, genomic medicine is still developing. For example, reports

have indicated that the detection of gene mutations in cancer patients is not always linked to the selection of therapeutic agents, so patient outcomes have not improved as much as we expected in this NGS era [29]. Even if this is the case, the importance of genomic medicine is expected to increase, along with the accumulation of genomic information that is closely associated with the development of molecular target drugs [29, 30].

In addition, for genomic analysis by NGS, no standard bioinformatics pipeline yet exists in which the entire analytical process is automated and qualified for clinical use, although the sensitivity and accuracy of detecting gene mutation shave clearly improved. This situation occurs at least partially because a wide variety of gene panels exist that depend on their purpose. Data characteristics differ for amplicon sequencing and capture-based panel sequencing. For example, the variant allele frequency might be overestimated or underestimated by a biased amplification of input DNA molecule in amplicon sequencing, while this can be largely prevented by capture-based sequencing because both ends of each fragment in a library prepared from the captured DNA are usually different. Also, panel-specific characteristics exist because of a high degree of flexibility in panel design, where a mutation detected by a gene panel cannot be always identified by another gene panel [31]. Therefore, a quite careful inspection of the clinical utility and of the validity of a designed gene panel is needed before clinical application.

Given this background, in this study we validated our original gene panel that targets 23 genes associated with hereditary tumors by analyzing peripheral blood samples obtained from patients who had previously been diagnosed as having the diseases. As a result, our gene panel successfully identified all mutations that were previously identified by means of Sanger sequencing. Our gene panel also newly identified a germline *APC* gene deletion and mutation in a patient who had a clinical diagnosis of FAP without having genetic tests. The position of the *APC* gene deletion in this patient was at exon 16, which suggested that the patient had AFAP; this conclusion was compatible with the phenotypic characteristics of the patient's polyps [25, 26]. Also, the infrequent variant p.Ser1126Arg in the *APC* gene was detected in this AFAP patient, although whether this variant and the frameshift mutation are in cis or in trans position remains to be determined. This missense variant, which was previously reported in Taiwanese patients with FAP [32, 33], is located close to the first 15-amino acid β-catenin binding repeat (amino acids 1136–1151) at the middle of the β-catenin downregulation domain, which may lead to disturbance in the putative cell signaling of the APC protein. Similarly, we changed the diagnosis for one of our patients with pheochromocytoma from HPPS (by an *SDHD* mutation) to VHL because of our identification of a pathogenic *VHL* gene mutation by NGS analysis. As mentioned in the Results section, inasmuch as an *SDHD* gene mutation was previously identified by using Sanger sequencing for this patient, other

genes including *VHL* had not been evaluated. If all genes that are involved in pheochromocytoma development had been tested at that time, the patient might not have been diagnosed with HPPS. However, examining the full set of multiple genes by Sanger sequencing simultaneously was not generally practical because of limited cost and time. Particularly in a case in which the restrictive order of genes to be tested is not fixed, a serial single-gene analysis will likely cause misinterpretation because of the overestimation of the first mutation identified. In view of these data, the NGS platform is superior to conventional Sanger sequencing, because NGS-based gene panel analysis can prevent overlooking such existing mutations.

Each NGS panel has its own characteristics, and sequencing data are not always uniformly available for all designed regions. In this study, we performed a gene panel analysis for hereditary tumors by means of the amplicon sequencing method, but a 614-bp gap existed when the panel was designed. Indeed, in our panel, an adequate sequencing depth, which was defined as having a depth of ≥30, was not achieved for an average of 8.1% across genes for all samples that we analyzed. Therefore, in our panel, there may be mutations that cannot be identified even though they exist, which would give rise to false-negative results. Combined application of Sanger sequencing or other methods with NGS for such regions is ideal but not always practical for all analyses. Thus, we should design a gene panel with as high uniformity as possible and re-design it if necessary.

In addition, careful interpretation of the pathogenic significance of identified mutations is clearly required. NGS-based panel sequencing scans large genome regions, so usually many non-pathogenic, benign gene variants will be detected. The ClinVar database, which we used as a reference in this study, contains about 41,800 variants with conflicting interpretations in about 794,000 records, and these variants are now reported as VUS. However, interpretations of these variants may change as data accumulate in the future. Similar to *RET* c.2071G>A in UPN8 and *MEN1* c.1636G>A in UPN10 that we detected in this study, certain gene variants were initially regarded as disease-causing variants but were later excluded because of relatively high allele frequencies. These misinterpretations were unavoidable because of the limited information available from the database when these variants were initially identified by Sanger sequencing, but as a result, the original diagnoses of the diseases were incorrect. Likewise, interpretations of variants that were judged to be VUS in this NGS-based analysis may be altered in the future. A periodic review of the identified VUS in the latest database is thus absolutely necessary.

Finally, in this study, we recognized the importance of manual inspection of NGS data. In the NGS system in which short fragments are amplified and aligned to reference genome databases, amplicons are not always amplified from the desired genome region. Actually, in this study, we found cases in which parts of *PMS2CL* or *PMS2P3* not

included in this panel were mis-amplified and aligned to *PMS2*. Inasmuch as these examples cannot be confirmed as true mutations by simply looking at the VCF file, ensuring the quality of the sequencing at the post-analytical phase by references to pre-existing data for samples, including those from normal individuals, obtained via the same gene panel, as well as by careful manual monitoring of the data by using a genome browser, are important.

In summary, validation of our original gene panel for hereditary tumors revealed the usefulness of our amplicon-based custom gene panel, as well as the need for skills to interpret the data produced by NGS.

Supplementary information is available in the online version of the paper.

References

[1] YU TM, MORRISON C, GOLD EJ, TRADONSKY A, LAYTON AJ. Multiple biomarker testing tissue consumption and completion rates with single-gene tests and investigational use of Oncomine Dx Target Test for advanced non-small-cell lung cancer: a single-center analysis. Clin Lung Cancer 2019; 20: 20–29. https://doi.org/10.1016/j.cllc.2018.08.010

[2] GOLDBERG KB, BLUMENTHAL GM, PAZDUR R. The first year of the Food and Drug Administration Oncology Center of Excellence: landmark approvals in a dynamic regulatory environment. Cancer J 2018; 24: 131–135. https://doi.org/10.1097/PPO.0000000000000316

[3] BAYSAL BE, FERRELL RE, WILLETT-BROZICK JE, LAWRENCE EC, MYSSIOREK D et al. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science 2000; 287: 848–851. https://doi.org/10.1126/science.287.5454.848

[4] GILL AJ. Succinate dehydrogenase (SDH)-deficient neoplasia. Histopathology 2018; 72: 106–116. https://doi.org/10.1111/his.13277

[5] LATIF F, TORY K, GNARRA J, YAO M, DUH FM et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. Science 1993; 260: 1317–1320. https://doi.org/10.1126/science.8493574

[6] CRESPIGIO J, BERBEL LCL, DIAS MA, BERBEL RF, PEREIRA SS et al. Von Hippel-Lindau disease: a single gene, several hereditary tumors. J Endocrinol Invest 2018; 41: 21–31. https://doi.org/10.1007/s40618-017-0683-1

[7] WIMMER K, WERNSTEDT A. PMS2 gene mutational analysis: direct cDNA sequencing to circumvent pseudogene interference. Methods Mol Biol 2014; 1167: 289–302. https://doi.org/10.1007/978-1-4939-0835-6_20

[8] CLENDENNING M, HAMPEL H, LAJEUNESSE J, LINDBLOM A, LOCKMAN J et al. Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 2006; 27: 490–495. https://doi.org/10.1002/humu.20318

[9] VAUGHN CP, HART KJ, SAMOWITZ WS, SWENSEN JJ. Avoidance of pseudogene interference in the detection of 3' deletions in PMS2. Hum Mutat 2011; 32: 1063–1071. https://doi.org/10.1002/humu.21540

[10] GREEN RC, BERG JS, GRODY WW, KALIA SS, KORF BR et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013; 15: 565–574. https://doi.org/10.1038/gim.2013.73

[11] HIGASA K, MIYAKE N, YOSHIMURA J, OKAMURA K, NIIHORI T et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. J Hum Genet 2016; 61: 547–553. https://doi.org/10.1038/jhg.2016.12

[12] ROBINSON JT, THORVALDSDOTTIR H, WINCKLER W, GUTTMAN M, LANDER ES et al. Integrative genomics viewer. Nat Biotechnol 2011; 29: 24–26. https://doi.org/10.1038/nbt.1754

[13] THORVALDSDOTTIR H, ROBINSON JT, MESIROV JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013; 14: 178–192. https://doi.org/10.1093/bib/bbs017

[14] JENNINGS LJ, ARCILA ME, CORLESS C, KAMEL-REID S, LUBIN LM et al. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn 2017; 19: 341–365. https://doi.org/10.1016/j.jmoldx.2017.01.011

[15] DE VOS M, HAYWARD BE, PICTON S, SHERIDAN E, BONTHRON DT. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet 2004; 74: 954–964. https://doi.org/10.1086/420796

[16] NAKAGAWA H, LOCKMAN JC, FRANKEL WL, HAMPEL H, STEENBLOCK K et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 2004; 64: 4721–4727. https://doi.org/10.1158/0008-5472.CAN-03-2879

[17] LANTIERI F, CAROLI F, CECCHERINI I, GRISERI P. The involvement of the RET variant G691S in medullary thyroid carcinoma enlightened by a meta-analysis study. Int J Cancer 2013; 132: 2808–2819. https://doi.org/10.1002/ijc.27967

[18] NORDSTROM-O'BRIEN M, VAN DER LUIJT RB, VAN ROOIJEN E, VAN DEN OUWELAND AM, MAJOOR-KRAKAUER DF et al. Genetic analysis of von Hippel-Lindau disease. Hum Mutat 2010; 31: 521–537. https://doi.org/10.1002/humu.21219

[19] BOEDEKER CC, ERLIC Z, RICHARD S, KONTNY U, GIMENEZ-ROQUEPLO AP et al. Head and neck paragangliomas in von Hippel-Lindau disease and multiple endocrine neoplasia type 2. J Clin Endocrinol Metab 2009; 94: 1938–1944. https://doi.org/10.1210/jc.2009-0354

[20] CRUZ JB, FERNANDES LP, CLARA SA, CONDE SJ, PER-
ONE D et al. Molecular analysis of the Von Hippel-Lindau
(VHL) gene in a family with non-syndromic pheochromo-
cytoma: the importance of genetic testing. Arq Bras Endocri-
nol Metabol 2007; 51: 1463–1467. https://doi.org/10.1590/
s0004-27302007000900008

[21] BAYSAL BE. Mitochondrial complex II and genomic im-
printing in inheritance of paraganglioma tumors. Biochim
Biophys Acta 2013; 1827: 573–577. https://doi.org/10.1016/j.
bbabio.2012.12.005

[22] FERGUSON-SMITH AC. Genomic imprinting: the emer-
gence of an epigenetic paradigm. Nat Rev Genet 2011; 12:
565–575. https://doi.org/10.1038/nrg3032

[23] INOUE A, JIANG L, LU F, ZHANG Y. Genomic imprinting
of Xist by maternal H3K27me3. Genes Dev 2017; 31: 1927–
1932. https://doi.org/10.1101/gad.304113.117

[24] NEUMANN HP, PAWLU C, PECZKOWSKA M, BAUSCH
B, MCWHINNEY SR et al. Distinct clinical features of para-
ganglioma syndromes associated with SDHB and SDHD
gene mutations. JAMA 2004; 292: 943–951. https://doi.
org/10.1001/jama.292.8.943

[25] RONCUCCI L, PEDRONI M, MARIANI F. Attenuated ad-
enomatous polyposis of the large bowel: present and future.
World J Gastroenterol 2017; 23: 4135–4139. https://doi.
org/10.3748/wjg.v23.i23.4135

[26] DE LEON MP, PEDRONI M, RONCUCCI L, DOMATI F,
ROSSI G et al. Attenuated polyposis of the large bowel: a
morphologic and molecular approach. Fam Cancer 2017; 16:
211–220. https://doi.org/10.1007/s10689-016-9938-9

[27] SORAVIA C, BERK T, MADLENSKY L, MITRI A, CHENG
H et al. Genotype-phenotype correlations in attenuated ad-
enomatous polyposis coli. Am J Hum Genet 1998; 62: 1290–
1301. https://doi.org/10.1086/301883

[28] QI XP, MA JM, DU ZF, YING RB, FEI J et al. RET germ-
line mutations identified by exome sequencing in a Chinese
multiple endocrine neoplasia type 2A/familial medullary
thyroid carcinoma family. PLoS One 2011; 6: e20353. https://
doi.org/10.1371/journal.pone.0020353

[29] LE TOURNEAU C, DELORD JP, GONCALVES A, GA-
VOILLE C, DUBOT C et al. Molecularly targeted therapy
based on tumour molecular profiling versus conventional
therapy for advanced cancer (SHIVA): a multicentre, open-
label, proof-of-concept, randomised, controlled phase 2 trial.
Lancet Oncol 2015; 16: 1324–1334. https://doi.org/10.1016/
S1470-2045(15)00188-6

[30] KURNIT KC, DUMBRAVA EEI, LITZENBURGER B,
KHOTSKAYA YB, JOHNSON AM et al. Precision oncol-
ogy decision support: current approaches and strategies for
the future. Clin Cancer Res 2018; 24: 2719–2731. https://doi.
org/10.1158/1078-0432.CCR-17-2494

[31] FUJIKI R, IKEDA M, YOSHIDA A, AKIKO M, YAO Y,
et al. Assessing the accuracy of variant detection in cost-
effective gene panel testing by next-generation sequencing.
J Mol Diagn 2018; 20: 572–582. https://doi.org/10.1016/j.
jmoldx.2018.04.004

[32] CHIANG JM, CHEN HW, TANG RP, CHEN JS,
CHANGCHIEN CR et al. Mutation analysis of the APC
gene in Taiwanese FAP families: low incidence of APC
germline mutation in a distinct subgroup of FAP families.
Fam Cancer 2010; 9: 117–124. https://doi.org/10.1007/
s10689-009-9292-2

[33] CHEN SP, TSAI ST, JAO SW, HUANG YL, CHAO YC et
al. Single nucleotide polymorphisms of the APC gene and
colorectal cancer risk: a case-control study in Taiwan. BMC
Cancer 2006; 6: 83. https://doi.org/10.1186/1471-2407-6-83