GMR

The Original

Target-enrichment capturing Next-Generation Sequencing panel for screening sarcoma associated driver mutations

JW Ahn[1]*, JH Kim[2]*, BR Kim[1], JH Min[1], HG Kang[2], SJ Kim[2], T Yun[2], WY Choi[2], YG Suh[2], SY Park[2], HJ You[2], EY Lee[2], YN Kim[2], JY Sung[2] and KW Kim[2]

[1]Celemics, Inc, 19-20F, BYC Highcity Bldg. A, Gasandigital 1-ro 131,Geumcheon-gu, Seoul, Korea.
[2]National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang-si Gyeonggi-do, 10408, Republic of Korea.

Corresponding author: JW Ahn
E-mail: jwahn@celemics.com

Corresponding author: JH Kim
E-mail: docjune@ncc.re.kr

Genet. Mol. Res. 23 (4): gmr2383
Received September 18, 2024
Accepted October 23, 2024
Published October 28, 2024
DOI http://dx.doi.org/10.4238/gmr2383

ABSTRACT. HSarcomas comprise over 70 subtypes of mesenchymal cancers. However, driving mutation remains still unclear for most of sarcoma types. Therefore, it is very important to secure data on driving mutations, it is necessary to incorporate Next-generation sequencing technology that analyzes multiple genes simultaneously.
In this article, we attempted to acheive an NGS panel that could capture genes in which sarcoma DNA mutations are frequently found. By using designed sarcoma NGS panel, we confirmed the applicability of clinical samples, confirmed the detection sensitivity of mutation frequency, and confirmed applicable sample types.

The result of testing samples with self-produced low frequency mutations using this panel, the limit of detection was found to be 0.2%. By applying to clinical samples, the performance of the panel was 60% on-target and the inter-probe capture variance value were average 0.34, confirming excellent balance between the sequenced regions.
Based on these results, our designed NGS panel is expected that the driving mutation can be detected from germline mutation to criculating tumor DNA level for the various sample type of sarcoma specimen, and it was found to be suitable for application to clinical specimens.
The spectrum of driving mutations is very important information in determining the direction of treatment for sarcoma patients, activities to accumulate data on the mutation spectrum are very necessary. As the spectrum of specific DNA driving mutations may differ between races depending on the cancer type, our selected genes targeting NGS panel will be useful in studying the spectrum of mutations specific to Koreans in cost-effective way.

INTRODUCTION

The Next-generation sequencing (NGS) technique has enabled to sequence the gene in various species. Discoveries of the genetic information, especially from coding regions, had a significant impact on the understanding of areas such as the onset of diseases, and have further contributed to various fields such as the development of treatments (Qin, 2019).

By taking advantage of these aspects, we have been able to build a database for diseases that are a priority of interest in the medical field, which are highly related to human life (Barker, 2022). In the case of cancer, NGS is used for customized treatment (Morganti et al., 2019), such as identifying the spectrum of driving mutations in most cancer types and customizing treatments for each main driving mutation, etc., thereby contributing greatly to extending the life of cancer patients. For this reason, clinical institutions are building customized NGS panels for various cancer type and using them clinically for diagnostic purposes (Chevrier et al., 2022; Dorwal et al., 2023; Kaderbhai et al., 2016; Suh et al., 2022; Surrey et al., 2019).

However, some of cancer types are difficult to construct an NGS panel for several reasons, such as difficulty in identifying clear driving gene or insufficient populational databases of some cancer types. Although regarded as rare carcinomas, they are reported to account for about 20% of all carcinomas (Keat et al., 2013), occupy a fairly high portion and are clinically important. Sarcomas are mesenchymal originated tumors, arising in connective tissue with rare incidence of less than 1% of all new cancer diagnoses (Kang, 2015). And approximately patients diagnosed as sarcoma annually in America is 3600 (Siegel, 2020), 4000 in Europe (Gatta et al., 2017), and 800 in Japan (Ogura, 2017). In Korea, Sarcoma is a representative rare type of cancer with a poor prognosis(Kang, 2015), with no clear genetic cause and therefore no clear treatment method and annual incidence is known as 760–1175/year (Kim et al., 2019). Sarcoma is known to have more than 70 subtypes, and these various subtypes contribute to a poor prognosis. For this reason, customizing a NGS panel that can be applied to sarcoma cancer patients and expand the capability of designed NGS panel test to clinical samples have great clinical and technical significance.

In this study, we collected information on genes related to the development of sarcoma and developed an NGS panel related to DNA mutations. We designed a hybridization capture-based NGS panel focused on SNV/indel variants of 129 genes in sarcoma, and evaluated the performance using the Illumina NGS sequencer platform.

MATERIALS AND METHODS

Selection of sarcoma panel genes and design of panel probes

Genes were selected by referring to the list of sarcoma genes (Supplementary Table S1) were found in the clinical database portal (TCGA; The Cancer Gene Atlas, Clinvar), sarcoma associated syndromes (Li-Fraumeni Syndrome, Retinoblastoma, Rothmund Thomson Syndrome, Werner Syndrome, Bloom Syndrome, Diamond Blackfan Anemia) and driving gene research papers (Boddu et al., 2018; Carmagnani Pestana, 2019; Helman, 2003; Savage, 2011; Taylor et al., 2011; Wu et al., 2020; Zhou et al., 2020) related to sarcoma, and the composition of the sarcoma cancer panel being used in clinical institutions in other countries.

Among the selected genes, we initially designed probe sequence consisted of 100-bp target regions with 50 bp of overlap (2x tiling density). We removed candidate probes meeting the exclusion criteria: includes redundancy region of the genome, homopolymers (more than eight consecutive identical bases), or extreme GC content (<10% or >90%).

For PCR amplification and transcription during the production process of the sarcoma panel as a RNA probes, pair of 20bp sequence were selected by the computational analysis. This pair of sequence were expected not to cause interference with the probes as similarity with probes sequence were very low, and added to each flanking site of probe design for universal probe amplification and transcription.

Synthesis of sarcoma panel probes from microchip oligo pool

Using a microchip oligo pool synthesis, oligonucleotides covering 534,469 bp were synthesized (Supplementary Table S2), and the chip pool contained 8.12 μg of single stranded oligonucleotide in 80 μl (GeneScript. USA). A total of 0.5 μl of chip pool, 10 μl of KAPA HiFi polymerase (KAPA BIOSYSTEMS, USA), 8 μl of dH_2O with 1 μl of each forward and reverse primer were used for probe amplification with the following PCR conditions: 98°C for 3 min; 12 cycles of 30 sec at 98°C, 30 sec at 60°C, and 30 sec at 72°C; and 10 min at 72°C. PCR amplified samples were purified with magnetic beads (Celemics Inc, Korea). Amplified probes were then transcribed into RNA probe pool by using AmpliScribe T7-Flash Transcription Kits (Biosearch Technologies, USA). Quantification were performend by TapeStation4200 (Agilent, USA) along with RNA ScreenTape reagents (Agilent, USA) and diluted into working concentration.

Target capture sequencing performance analysis using sarcoma panel

HapMap samples (NA12878, NA12891 and NA12892; Coriell, USA) and OncoSpan FFPE DNA (HD832; Horizon, USA) were used as reference samples. Sarcoma panel performance tests were conducted depending on the amount of sample, from 25ng to 50ng, 75ng, and 100ng.

In the process of adjusting the allele frequency of the mutation to test sensitivity, test samples with various allele frequencies were produced by mixing NA12892 among the FFPE reference sample and HapMap sample at a certain ratio. The allele frequency analysis section was selected according to the frequency at which each mutation type (germline, somatic, ctDNA) typically occurs. The allele frequency analysis section was divided into a germline analysis section of more than 10%, a somatic mutation allele frequency section from 10% to 1%, and finally, a cfDNA mutation allele frequency section of less than 1%, and tests were conducted down to a minimum of 0.1%.

DNA from reference sample were fragmented, adapter ligated with the Illumina adapter, amplified with Illumina unique dual index primer pair using Celemics EP-Kit (Celemics Inc, Korea). Sarcoma panel (Celemics Inc, Korea) were used to hybridize 500ng of Illumina index-tagged library according to Celemics target capture protocol. Target capture libraries were sequenced with the Illumina NextSeq550 platform (Illumina, USA) with 2×150-bp paired-end runs.

Target capture and NGS sequencing of samples

To confirm performance in actual clinical samples, we prospectively collected clinical tumor samples in patients with sarcoma. This study was approved by the Institutional Review Board of National Cancer Center of Korea (IRB No. NCC2021-0184), and all patients provided informed consent for tissue banking and genetic testing. Since collected bone sarcoma samples were attached to the bone, it was shredded using a TissueRuptor II (Qiagen ,USA) to ease the DNA extraction step. DNA were extracted by the DNeasy Blood & Tissue Kit (Qiagen, USA) and extracted gDNA were quantified and qualified using TapeStation4200 with Genomic DNA ScreenTape reagents (Agilent, USA). Target gene capture using the sarcoma panel was carried out through the same process as the performance test using the reference sample, and 100ng of input DNA was used.

Bioinformatics and data analysis for target captured NGS date

Raw FASTQ file was filtered and trimmed using Adaptor removal 2.3.3. Burrows-Wheeler aligner (BWA; version 0.7.17) mem with default option was used to align reads to human reference genome sequence GRCh37. Sequence alignment map (SAM) file was converted to BAM format using samtools (version 1.15.1). Picard tool (version 2.27.2) was used to sort and remove duplications. GATK (version 4.2.6.1) was used to perform Base quality score re-calibration and variant calling. GATK HaplotypeCaller was used to call variants following with the parameters -stand_call_conf 50 -stand_emit_conf 30. Format converting, sorting, and indexing were performed by Samtools. Detected variants were annotated with ANNOVAR. False positive indels were removed by manually reviewing each indel with the Intergrated Genome Viewer (IGV version 1.8.0.).

RESULTS

Characteristics of sarcoma gene screening panel

In the present study, we developed gene screening panel for sarcoma (here on, referred as sarcoma panel) using next-generation sequencing (NGS) along with in-house program for bioinformatic linking for each variants. Sarcoma panel was designed to detects single nucleotide variants (SNVs) and small insertions/deletions (Indel; up to 25 bp) located in the DNA coding sequences of 129 genes (Figure 1), nearby flanking regions (20 bp flanking of each exon) and known splice regions in the targeted genes.

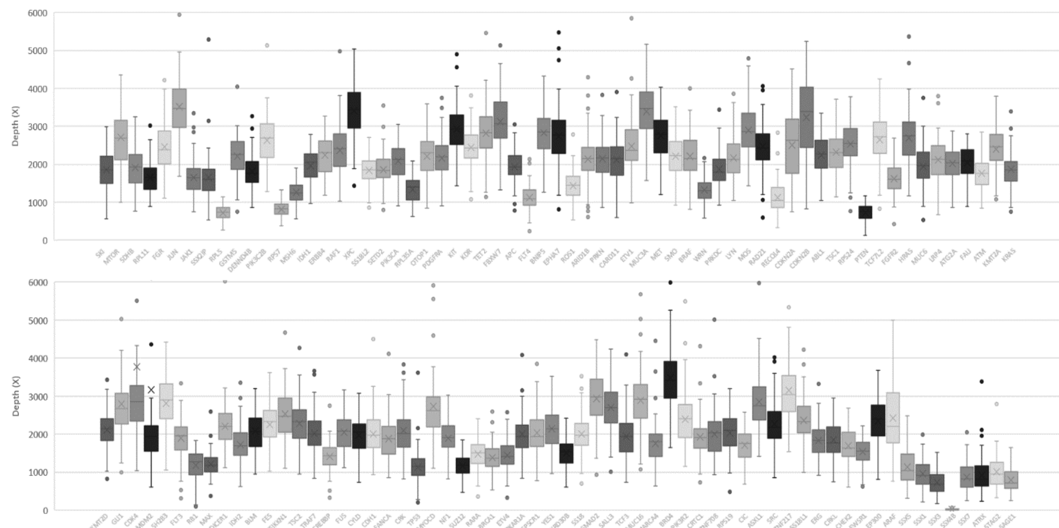


Figure 1. Overall depth status with designed target region from 128 genes of sarcoma panel.

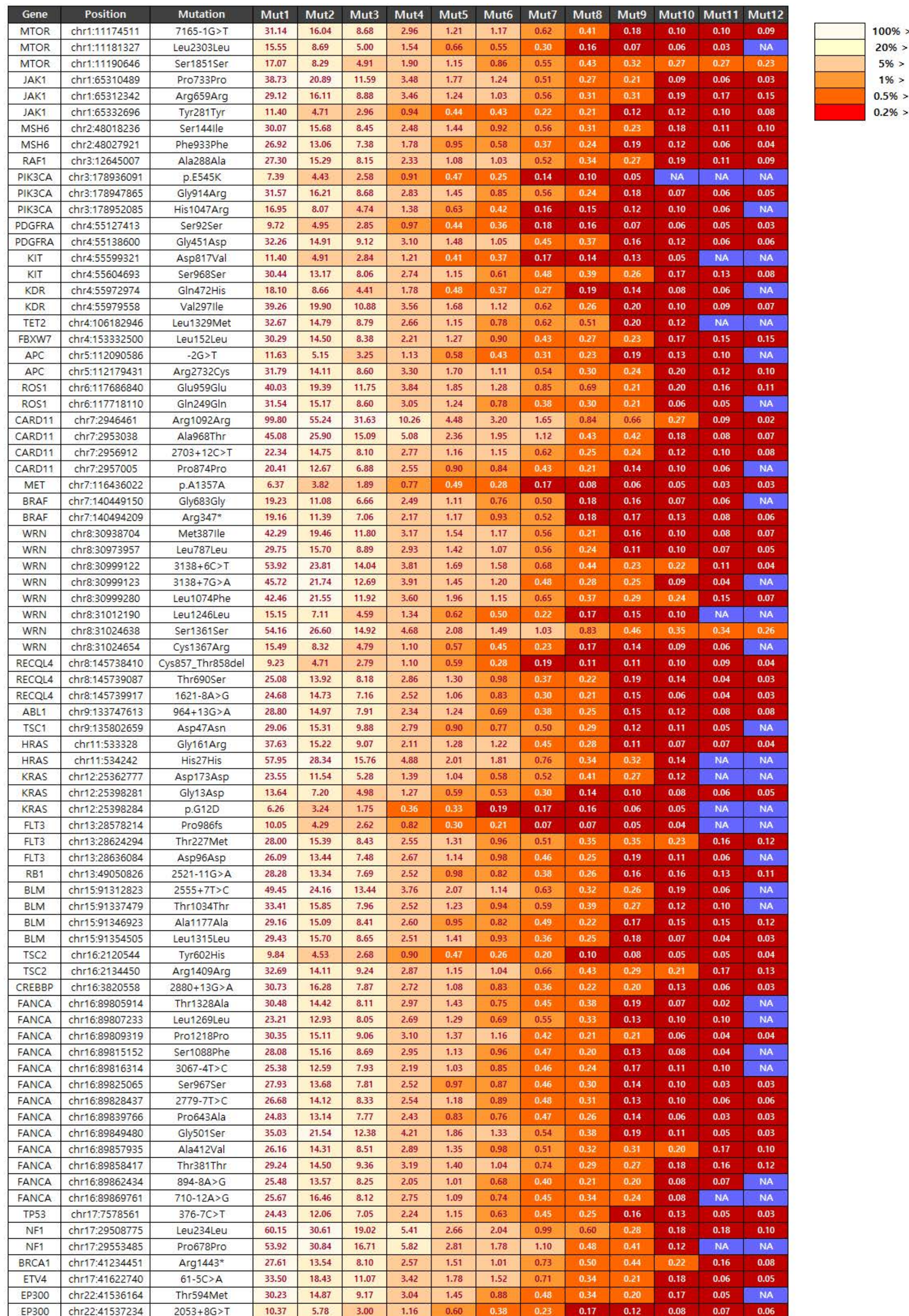
The sarcoma panel was generated based on RNA probes, which can be mass-produced using transcription from sarcoma panel probes composed of DNA, which can reduce the cost burden of clinical testing when used for clinical purposes in the future and the strong RNA:DNA complex which have advantage in high recovery rate of the target gene region using binding affinity.

Sarcoma panel performance and variant detection with reference materials

The basic target gene coverage and capture uniformity between probes were analyzed by obtaining data from the HapMap reference sample. On-target reads ratio were average 62.8% and internal uniformity among probes (standard deviation of probes) were 0.34. Based on this data we have moved to next-step using reference DNA from FFPE materials. Overall exonic region of the target genes were confirmed uniform depth was achieved (Figure 1) with average un-covered region 0.79% (Supplementary Table S3).

Next, first we compared the impact of different starting amount of DNA on sarcoma panel performance. And we confirmed that there was no difference in detection sensitivity and sarcoma panel capture performance for known variants in all conditions, from the lowest input DNA condition tested, 25ng, to the highest input DNA condition, 100ng (average on-target reads ratio %, probe uniformity).

Second, we tested the difference in panel performance depending on the presence or absence of interfering factors that may be introduced during the process of securing sarcoma DNA samples. In this process, DNA damage caused by heat generated during the bone grinding process to extract DNA from bone samples, deparaffinization solution that can be included when extracting DNA from the FFPE block, ethanol that can be included in the DNA purification process, and hemoglobin that can be contained from the blood sample were included. Overall, no deterioration in sarcoma panel performance was observed due to these external factor (average on-target reads ratio %, probe uniformity).



Gene	Position	Mutation	Mut1	Mut2	Mut3	Mut4	Mut5	Mut6	Mut7	Mut8	Mut9	Mut10	Mut11	Mut12
MTOR	chr1:11174511	7165-1G>T	31.14	16.04	8.68	2.96	1.21	1.17	0.62	0.41	0.18	0.10	0.10	0.09
MTOR	chr1:11181327	Leu2303Leu	15.55	8.69	5.00	1.54	0.66	0.55	0.30	0.16	0.07	0.06	0.03	NA
MTOR	chr1:11190646	Ser1851Ser	17.07	8.29	4.91	1.90	1.15	0.86	0.55	0.43	0.32	0.27	0.27	0.23
JAK1	chr1:65310489	Pro733Pro	38.73	20.89	11.59	3.48	1.77	1.24	0.51	0.27	0.21	0.09	0.06	0.03
JAK1	chr1:65312342	Arg659Arg	29.12	16.11	8.88	3.46	1.24	1.03	0.56	0.31	0.31	0.19	0.17	0.15
JAK1	chr1:65332696	Tyr281Tyr	11.40	4.71	2.96	0.94	0.44	0.43	0.22	0.21	0.12	0.12	0.10	0.08
MSH6	chr2:48018236	Ser144Ile	30.07	15.68	8.45	2.48	1.44	0.92	0.56	0.31	0.23	0.18	0.11	0.10
MSH6	chr2:48027921	Phe933Phe	26.92	13.06	7.38	1.78	0.95	0.58	0.37	0.24	0.19	0.12	0.06	0.04
RAF1	chr3:12645007	Ala288Ala	27.30	15.29	8.15	2.33	1.08	1.03	0.52	0.34	0.27	0.19	0.11	0.09
PIK3CA	chr3:178936091	p.E545K	7.39	4.43	2.58	0.91	0.47	0.25	0.14	0.10	0.05	NA	NA	NA
PIK3CA	chr3:178947865	Gly914Arg	31.57	16.21	8.68	2.83	1.45	0.85	0.56	0.24	0.18	0.07	0.06	0.05
PIK3CA	chr3:178952085	His1047Arg	16.95	8.07	4.74	1.38	0.63	0.42	0.16	0.15	0.12	0.10	0.06	NA
PDGFRA	chr4:55127413	Ser92Ser	9.72	4.95	2.85	0.97	0.44	0.36	0.18	0.16	0.07	0.06	0.05	0.03
PDGFRA	chr4:55138600	Gly451Asp	32.26	14.91	9.12	3.10	1.48	1.05	0.45	0.37	0.16	0.12	0.06	0.06
KIT	chr4:55599321	Asp817Val	11.40	4.91	2.84	1.21	0.41	0.37	0.17	0.14	0.13	0.05	NA	NA
KIT	chr4:55604693	Ser968Ser	30.44	13.17	8.06	2.74	1.15	0.61	0.48	0.39	0.26	0.17	0.13	0.08
KDR	chr4:55972974	Gln472His	18.10	8.66	4.41	1.78	0.48	0.37	0.27	0.19	0.14	0.08	0.06	NA
KDR	chr4:55979558	Val297Ile	39.26	19.90	10.88	3.56	1.68	1.12	0.62	0.26	0.20	0.10	0.09	0.07
TET2	chr4:106182946	Leu1329Met	32.67	14.79	8.79	2.66	1.15	0.78	0.62	0.51	0.20	0.12	NA	NA
FBXW7	chr4:153332500	Leu152Leu	30.29	14.50	8.38	2.21	1.27	0.90	0.43	0.27	0.23	0.17	0.15	0.15
APC	chr5:112090586	-2G>T	11.63	5.15	3.25	1.13	0.58	0.43	0.31	0.23	0.19	0.13	0.10	NA
APC	chr5:112179431	Arg2732Cys	31.79	14.11	8.60	3.30	1.70	1.11	0.54	0.30	0.24	0.20	0.12	0.10
ROS1	chr6:117686840	Glu959Glu	40.03	19.39	11.75	3.84	1.85	1.28	0.85	0.69	0.21	0.20	0.16	0.11
ROS1	chr6:117718110	Gln249Gln	31.54	15.17	8.60	3.05	1.24	0.78	0.38	0.30	0.21	0.06	0.05	NA
CARD11	chr7:2946461	Arg1092Arg	99.80	55.24	31.63	10.26	4.48	3.20	1.65	0.84	0.66	0.27	0.09	0.02
CARD11	chr7:2953038	Ala968Thr	45.08	25.90	15.09	5.08	2.36	1.95	1.12	0.43	0.42	0.18	0.08	0.07
CARD11	chr7:2956912	2703+12C>T	22.34	14.75	8.10	2.77	1.16	1.15	0.62	0.25	0.24	0.12	0.10	0.08
CARD11	chr7:2957005	Pro874Pro	20.41	12.67	6.88	2.55	0.90	0.84	0.43	0.21	0.14	0.10	0.06	NA
MET	chr7:116436022	p.A1357A	6.37	3.82	1.89	0.77	0.49	0.28	0.17	0.08	0.06	0.05	0.03	0.03
BRAF	chr7:140449150	Gly683Gly	19.23	11.08	6.66	2.49	1.11	0.76	0.50	0.18	0.16	0.07	0.06	NA
BRAF	chr7:140494209	Arg347*	19.16	11.39	7.06	2.17	1.17	0.93	0.52	0.18	0.17	0.13	0.08	0.06
WRN	chr8:30938704	Met387Ile	42.29	19.46	11.80	3.17	1.54	1.17	0.56	0.21	0.16	0.10	0.08	0.07
WRN	chr8:30973957	Leu787Leu	29.75	15.70	8.89	2.93	1.42	1.07	0.56	0.24	0.11	0.10	0.07	0.05
WRN	chr8:30999122	3138+6C>T	53.92	23.81	14.04	3.81	1.69	1.58	0.68	0.44	0.23	0.22	0.11	0.04
WRN	chr8:30999123	3138+7G>A	45.72	21.74	12.69	3.91	1.45	1.20	0.48	0.28	0.25	0.09	0.04	NA
WRN	chr8:30999280	Leu1074Phe	42.46	21.55	11.92	3.60	1.96	1.15	0.65	0.37	0.29	0.24	0.15	0.07
WRN	chr8:31012190	Leu1246Leu	15.15	7.11	4.59	1.34	0.62	0.50	0.22	0.17	0.15	0.10	NA	NA
WRN	chr8:31024638	Ser1361Ser	54.16	26.60	14.92	4.68	2.08	1.49	1.03	0.83	0.46	0.35	0.34	0.26
WRN	chr8:31024654	Cys1367Arg	15.49	8.32	4.79	1.10	0.57	0.45	0.23	0.17	0.14	0.09	0.06	NA
RECQL4	chr8:145738410	Cys857_Thr858del	9.23	4.71	2.79	1.10	0.59	0.28	0.19	0.11	0.11	0.10	0.09	0.04
RECQL4	chr8:145739087	Thr690Ser	25.08	13.92	8.18	2.86	1.30	0.98	0.37	0.22	0.19	0.14	0.04	0.03
RECQL4	chr8:145739917	1621-8A>G	24.68	14.73	7.16	2.52	1.06	0.83	0.30	0.21	0.15	0.06	0.04	0.03
ABL1	chr9:133747613	964+13G>A	28.80	14.97	7.91	2.34	1.24	0.69	0.38	0.25	0.15	0.12	0.08	0.08
TSC1	chr9:135802659	Asp47Asn	29.06	15.31	9.88	2.79	0.90	0.77	0.50	0.29	0.12	0.11	0.05	NA
HRAS	chr11:533328	Gly161Arg	37.63	15.22	9.07	2.11	1.28	1.22	0.45	0.28	0.11	0.07	0.07	0.04
HRAS	chr11:534242	His27His	57.95	28.34	15.76	4.88	2.01	1.81	0.76	0.34	0.32	0.14	NA	NA
KRAS	chr12:25362777	Asp173Asp	23.55	11.54	5.28	1.39	1.04	0.58	0.52	0.41	0.27	0.12	NA	NA
KRAS	chr12:25398281	Gly13Asp	13.64	7.20	4.98	1.27	0.59	0.53	0.30	0.14	0.10	0.08	0.06	0.05
KRAS	chr12:25398284	p.G12D	6.26	3.24	1.75	0.36	0.33	0.19	0.17	0.16	0.06	0.05	NA	NA
FLT3	chr13:28578214	Pro986fs	10.05	4.29	2.62	0.82	0.30	0.21	0.07	0.07	0.05	0.04	NA	NA
FLT3	chr13:28624294	Thr227Met	28.00	15.39	8.43	2.55	1.31	0.96	0.51	0.35	0.35	0.23	0.16	0.12
FLT3	chr13:28636084	Asp96Asp	26.09	13.44	7.48	2.67	1.14	0.98	0.46	0.25	0.19	0.11	0.06	NA
RB1	chr13:49050826	2521-11G>A	28.28	13.34	7.69	2.52	0.98	0.82	0.38	0.26	0.16	0.16	0.13	0.11
BLM	chr15:91312823	2555+7T>C	49.45	24.16	13.44	3.76	2.07	1.14	0.63	0.32	0.26	0.19	0.06	NA
BLM	chr15:91337479	Thr1034Thr	33.41	15.85	7.96	2.52	1.23	0.94	0.59	0.39	0.27	0.12	0.10	NA
BLM	chr15:91346923	Ala1177Ala	29.16	15.09	8.41	2.60	0.95	0.82	0.49	0.22	0.17	0.15	0.15	0.12
BLM	chr15:91354505	Leu1315Leu	29.43	15.70	8.65	2.51	1.41	0.93	0.36	0.25	0.18	0.07	0.04	0.03
TSC2	chr16:2120544	Tyr602His	9.84	4.53	2.68	0.90	0.47	0.26	0.20	0.10	0.08	0.05	0.05	0.04
TSC2	chr16:2134450	Arg1409Arg	32.69	14.11	9.24	2.87	1.15	1.04	0.66	0.43	0.29	0.21	0.17	0.13
CREBBP	chr16:3820558	2880+13G>A	30.73	16.28	7.87	2.72	1.08	0.83	0.36	0.22	0.20	0.13	0.06	0.03
FANCA	chr16:89805914	Thr1328Ala	30.48	14.42	8.11	2.97	1.43	0.75	0.45	0.38	0.19	0.07	0.02	NA
FANCA	chr16:89807233	Leu1269Leu	23.21	12.93	8.05	2.69	1.29	0.69	0.55	0.33	0.13	0.10	0.10	NA
FANCA	chr16:89809319	Pro1218Pro	30.35	15.11	9.06	3.10	1.37	1.16	0.42	0.21	0.21	0.06	0.04	0.04
FANCA	chr16:89815152	Ser1088Phe	28.08	15.16	8.69	2.95	1.13	0.96	0.47	0.20	0.13	0.08	0.04	NA
FANCA	chr16:89816314	3067-4T>C	25.38	12.59	7.93	2.19	1.03	0.85	0.46	0.24	0.17	0.11	0.10	NA
FANCA	chr16:89825065	Ser967Ser	27.93	13.68	7.81	2.52	0.97	0.87	0.46	0.30	0.14	0.10	0.03	0.03
FANCA	chr16:89828437	2779-7T>C	26.68	14.12	8.33	2.54	1.18	0.89	0.48	0.31	0.13	0.10	0.06	0.06
FANCA	chr16:89839766	Pro643Ala	24.83	13.14	7.77	2.43	0.83	0.76	0.47	0.26	0.14	0.06	0.03	0.03
FANCA	chr16:89849480	Gly501Ser	35.03	21.54	12.38	4.21	1.86	1.33	0.54	0.38	0.19	0.11	0.05	0.03
FANCA	chr16:89857935	Ala412Val	26.16	14.31	8.51	2.89	1.35	0.98	0.51	0.32	0.31	0.20	0.17	0.10
FANCA	chr16:89858417	Thr381Thr	29.24	14.50	9.36	3.19	1.40	1.04	0.74	0.29	0.27	0.18	0.16	0.12
FANCA	chr16:89862434	894-8A>G	25.48	13.57	8.25	2.05	1.01	0.68	0.40	0.21	0.20	0.08	0.07	NA
FANCA	chr16:89869761	710-12A>G	25.67	16.46	8.12	2.75	1.09	0.74	0.45	0.34	0.24	0.08	NA	NA
TP53	chr17:7578561	376-7C>T	24.43	12.06	7.05	2.24	1.15	0.63	0.45	0.25	0.16	0.13	0.05	0.03
NF1	chr17:29508775	Leu234Leu	60.15	30.61	19.02	5.41	2.66	2.04	0.99	0.60	0.28	0.18	0.18	0.10
NF1	chr17:29553485	Pro678Pro	53.92	30.84	16.71	5.82	2.81	1.78	1.10	0.48	0.41	0.12	NA	NA
BRCA1	chr17:41234451	Arg1443*	27.61	13.54	8.10	2.57	1.51	1.01	0.73	0.50	0.44	0.22	0.16	0.08
ETV4	chr17:41622740	61-5C>A	33.50	18.43	11.07	3.42	1.78	1.52	0.71	0.34	0.21	0.18	0.06	0.05
EP300	chr22:41536164	Thr594Met	30.23	14.87	9.17	3.04	1.45	0.88	0.48	0.34	0.20	0.17	0.05	NA
EP300	chr22:41537234	2053+8G>T	10.37	5.78	3.00	1.16	0.60	0.38	0.23	0.17	0.12	0.08	0.07	0.06

Figure 2. Detection sensitivity test result in 80 mutations from reference sample. The value in box is the detected frequency percentage (%) of the mutation.

Lastly, the detection sensitivity was compared according to the allele frequency of the mutation in 100ng input of DNA. In the sensitivity test, considering the field situation such as the rarity of sarcoma samples, the input DNA conditions that seemed to have the highest possibility of actual use were set and analysis was performed with 100ng. The detection sensitivity was confirmed to be possible down to 0.2% or less, and the detection sensitivity was confirmed through a repeatability test (Figure 1). Through this process, it was confirmed that all mutations contained in the sample were detected in all repeated experiments for conditions higher than 0.1%. Based on this data, the limit of detection of the test panel was set to 0.2% (Figure 2). We found that the concordance between replicates of the intra and inter-run replicates was >99.9%. Through this process, it was determined that the usability of the sarcoma panel was worth applying to clinical samples, so we attempted to confirm its performance by applying it to clinical samples.

Sarcoma panel performance on clinical sarcoma specimens

We have analyzed sequencing data from 88 clinical samples to compare performance with reference sample. Sequencing data generated mean of 27 million reads per sample (range 12~37 million) to achieve average of 1000X depth or more for condition to detect variants of less than 1%. After PCR duplicate reads removal, on-target reads on target genomic regions were 60.4% ~80.4% (average 65.1%) with probe uniformity 0.43. All samples had more than 99% mapping rate and average 1,000X depth over covered region were obtained. Less than 4% of targeted regions were not covered, average 99.1% of region were covered with 100X depth. Assessing the exonic coverage in the targeted region, average of coverage of uncovered region of ovear all 2,359 exons were 0.62% (Supplementary Table S4).

DISCUSSION

As with TCGA, the overall gene mutation frequency can be confirmed by collecting a large number of samples for each cancer type, and the causative genes and corresponding treatments are known for most cancer types.

However, in the case of some rare carcinomas, including sarcoma, genetic statistics are still unclear due to the rarity of specimen collection, major driving mutations are not identified, and mutations in suppressor genes are mostly identified. These kind of huddle make difficulties in discovering the exact pathogenesis mechanism or treatment in Sarcoma. In particular, in the case of Korea, databases for rare cancer type like sarcoma are not yet well established. Even if there are some genetic similarities between races, there are significant statistical differences in detailed data between races, so it is important to build a country-specific race disease database for accurate research.

For this purpose, although the cost of sequencing has become cheaper than NGS were first introduced, in order to use it clinically or actively for research purposes, a gene panel consisting of only selected genes is needed to lower the entry barrier for research on rare cancers and improves data acquisition for large scale.

To address these challenges, we have developed and validated a 129-gene sarcoma panel using Illumina's NGS platform. This panel is designed to sequence and analyze the protein-coding regions of the targeted genes for the identification of sarcoma-causing mutations. The aim of study was to assess the analytical performances of the panel. No false positive was identified in the analytical

validation study. Also, it was designed and verified to be able to detect mutations at various levels, including germline mutations, somatic mutations, and ctDNA mutations, as explained in the results section. Although research on ctDNA in the sarcoma field is not yet active, this is an area that can be usefully applied when research in the sarcoma field expands in the future.

Currently, because sarcoma samples are rare compared to other cancer types and sample collection is difficult, the genes with statistical significance for each detailed subtype are not clear, making it difficult to build a Sarcoma panel for each detailed subtype. As data accumulates and genetic significance becomes clear in subtypes of sarcoma, it is expected that organizing a panel by detailed subtype through updating the panel will be effective in the clinical field. Through multi-institutional joint research such as the Biobank Innovation Consortium for Sarcoma Research (BICSWAN) could accelerate accumulation of clinical genetic data by subtype more quickly.

However, as described above, due to the fact that the spectrum of genes related to the development of sarcoma is mainly observed in tumor suppressor genes such as TP53, rather than designing a more focused panel, a panel with expanded versatility was produced. In the future, sarcoma panel gene condensing process may be necessary based on accumulated sequence data on sarcoma. In particular, with 60% on-target and optimized probe distribution configuration, it is expected that efficient sequencing will be possible with a depth difference between probes of 0.34, and it is expected to contribute to the study of sarcoma in Koreans.

In summary, we developed a NGS panel with a focus on clinically actionable mutations and validated the performance in library construction, sequencing and variant calling. Our results illustrate the feasibility for sarcoma panel, suitable for clinical application, improving diagnosis, prognosis and personalized therapeutic decisions.

CONCLUSION

Through this study, we developed an NGS panel consisting of 129 genes which is designed to detect clinically relevant DNA mutations associated with Sarcoma cancer types. The performance of the test is assessed across a board range of variants in the 129 genes to support clinical use. Developed sarcoma panel was also applied to clinical specimens to confirm that it can be sufficiently utilized for accurate detection of mutations, and it was confirmed that it has the ability to detect mutations at the germline, somatic, and even ctDNA levels.

Acknowledgement

This study was conducted using biospecimens and data from the Sarcoma Biobank Innovation Consortium for Biomedical and Healthcare Research, which was funded by the Korea Disease Control and Prevention Agency for the Korea Biobank Project (#6637-303).

REFERENCES

Barker AD, Lee JSH. (2022). Translating "Big Data" in Oncology for Clinical Benefit: Progress or Paralysis. *Cancer Res.* 82: 2072-2075.

Boddu S, Walko CM, Bienasz S, Bui MM, et al. (2018). Clinical Utility of Genomic Profiling in the Treatment of Advanced Sarcomas: A Single-Center Experience. *JCO Precis Oncol.* 2: 1-8.

Carmagnani Pestana R, Groisberg R, Roszik J, Subbiah V. (2019). Precision Oncology in Sarcomas: Divide and Conquer. *JCO Precis Oncol.* 3.

Chevrier S, Brasselet A, Carnet M, Chevriaux A, et al. (2022). Custom multi-tumor next-generation sequencing panel for routine molecular diagnosis of solid tumors: Validation and results from three-year clinical use. *Int J Mol Med.* 49.

Dorwal P, Gray A, Rathi V, Chan Y. (2023). Primary common bile duct melanoma: NGS based fusion panel helps in differentiation from gastrointestinal clear cell sarcoma. *Pathology.* 55: 423-425.

Gatta G, Capocaccia R, Botta L, Mallone S, et al. (2017). Burden and centralised treatment in Europe of rare tumours: results of RARECAREnet-a population-based study. *Lancet Oncol.* 18: 1022-1039.

Helman LJ, Meltzer P. (2003). Mechanisms of sarcoma development. *Nat Rev Cancer.* 3: 685-694.

Kaderbhai CG, Boidot R, Beltjens F, Chevrier S, et al. (2016). Use of dedicated gene panel sequencing using next generation sequencing to improve the personalized care of lung cancer. *Oncotarget.* 7: 24860-24870.

Kang S, Kim HS, Choi ES, Han I. (2015). Incidence and Treatment Pattern of Extremity Soft Tissue Sarcoma in Korea, 2009-2011: A Nationwide Study Based on the Health Insurance Review and Assessment Service Database. *Cancer Res Treat.* 47: 575-582.

Keat N, Law K, Seymour M, Welch J, et al. (2013). International rare cancers initiative. *Lancet Oncol.* 14: 109-110.

Kim HS, Nam CM, Jang SY, Choi SK, et al. (2019). Characteristics and Treatment Patterns of Patients with Advanced Soft Tissue Sarcoma in Korea. *Cancer Res Treat.* 51: 1380-1391.

Morganti S, Tarantino P, Ferraro E, D'Amico P, et al. (2019). Next Generation Sequencing (NGS): A Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer. *Adv Exp Med Biol.* 1168: 9-30.

Ogura K, Higashi T, Kawai A. (2017). Statistics of bone sarcoma in Japan: Report from the Bone and Soft Tissue Tumor Registry in Japan. *J Orthop Sci*. 22: 133-143.

Qin D. (2019). Next-generation sequencing and its clinical application. *Cancer Biol Med.* 16: 4-10.

Savage SA, Mirabello L. (2011). Using epidemiology and genomics to understand osteosarcoma etiology. *Sarcoma*. 2011: 548151.

Siegel RL, Miller KD, Jemal A. (2020). Cancer statistics, 2020. *CA Cancer J Clin.* 70: 7-30.

Suh KJ, Kim SH, Kim YJ, Shin H, et al. (2022). Clinical Application of Next-Generation Sequencing in Patients With Breast Cancer: Real-World Data. *J Breast Cancer*. 25: 366-378.

Surrey LF, MacFarland SP, Chang F, Cao K, et al. (2019). Clinical utility of custom-designed NGS panel testing in pediatric tumors. *Genome Med.* 11: 32.

Taylor BS, Barretina J, Maki RG, Antonescu CR, et al. (2011). Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer.* 11: 541-557.

Wu CC, Beird HC, Andrew Livingston J, Advani S, et al. (2020). Immuno-genomic landscape of osteosarcoma. *Nat Commun*. 11: 1008.

Zhou Y, Yang D, Yang Q, Lv X, et al. (2020). Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat Commun.* 11: 6322.