



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

의학박사 학위논문

**Whole-genome Sequencing Reveals  
Comprehensive Genomic Profiles of  
Radiation Induced Sarcomas**

전장 유전체 분석을 이용한  
방사선 유발 육종의 유전적 특성 규명

2021년 2월

서울대학교 대학원

임상의과학과

김 은 지

# Whole-genome Sequencing Reveals Comprehensive Genomic Profiles of Radiation Induced Sarcomas

지도교수 우 홍 균

이 논문을 의학박사 학위논문으로 제출함

2020년 10월






서울대학교 대학원

임상의과학과

김 은 지

김은지의 의학박사 학위논문을 인준함

2021년 1월

위 원 장	<u>김 인 아</u>	
부위원장	<u>우 홍 균</u>	
위 원	<u>한 일 귀</u>	
위 원	<u>문 경 현</u>	
위 원	<u>김 태 민</u>	

## Abstract

# Whole-genome Sequencing Reveals Comprehensive Genomic Profiles of Radiation Induced Sarcomas

Eunji Kim

*Department of Clinical Medical Sciences,  
Graduate School,  
Seoul National University*

**Background and Purpose:** Radiation-induced sarcoma (RIS) is a rare secondary malignancy that is caused by treatment-related ionizing radiation after a long latency period. Despite unfavorable clinical outcomes, the genomic footprints of ionizing radiations in RIS development remain largely unknown. Hence, this study was aimed at characterizing the genomes and analyzing the genomic alterations in RIS.

**Materials and Methods:** The patients with secondary sarcoma associated with radiotherapy were reviewed between 2000 and 2019. Thirty sarcomas developed in the previously irradiated area were reviewed by two experienced pathologists. DNA sample was extracted from freshly frozen tumor tissues or isolated tumor through microdissection, along with normal tissue or blood derived from the same individuals. We finally enrolled 11 samples for which libraries were successfully

created. Whole-genome sequencing (WGS) was performed with the average coverage of tumor 90 × and of normal 60 ×. The pipelines for analyzing single nucleotide variations, short insertion/deletion, somatic copy number alterations, structural variations, and germline mutations were constructed.

**Results:** The mutation abundance of RIS genomes including one hypermutated genome was variable. Cancer-related genes might show different types of genomic alterations. For instance, *NF1*, *NF2*, *NOTCH1*, *NOTCH2*, *PIK3CA*, *RBI*, and *TP53* showed singleton somatic mutations; *MYC*, *CDKN2A*, *RBI*, and *NF1* showed recurrent copy number alterations; and *NF2*, *ARID1B*, and *RAD51B* showed recurrent structural variations (SVs). The effects of non-homologous end joining on short insertions-deletions and SVs were substantial in RIS genomes, compared with in spontaneous osteosarcoma genomes, representing the genomic hallmark of RIS genomes. In addition, frequent chromothripsis and predisposing germline variants in DNA damage-repair pathways were identified.

**Conclusion:** Taken together, WGS-scale characterization of RIS genomes may pave the way for advanced diagnostic and therapeutic strategies for RIS.

**Keywords:** Whole genome sequencing, Radiation induced sarcoma, Radiation, Second malignancy, Somatic mutation, Germline mutation, Structural variation

**Student number:** 2015-30808

# Contents

<b>Abstract</b> .....	i
<b>Contents</b> .....	iii
<b>List of Figures</b> .....	iv
<b>List of Tables</b> .....	vi
<b>Introduction</b> .....	1
<b>Materials and Methods</b> .....	3
<b>Results</b> .....	8
<b>Discussion</b> .....	47
<b>Conclusion</b> .....	51
<b>References</b> .....	52
<b>Abstract in Korean</b> .....	61

## List of Figures

<b>Figure 1.</b>	Tumor mutation burdens (TMB) in RIS genomes .....	14
<b>Figure 2.</b>	Comparison of TMB between RIS and radiation- naïve sarcomas...	15
<b>Figure 3.</b>	Mutation features with respect to nearby (A) genes and (B) spectra in 11RIS genomes .....	17
<b>Figure 4.</b>	De novo discovery of mutation signatures in 11 RIS genomes .....	18
<b>Figure 5.</b>	Mutation signatures of SNV in 3 cohorts .....	20
<b>Figure 6.</b>	Mutation signatures of indel in 3 cohorts .....	21
<b>Figure 7.</b>	Comparison of COSMIC (A) SBS and (B) ID mutation signatures Between RIS and radiation- naïve sarcomas .....	22
<b>Figure 8.</b>	Somatic mutations observed in cancer-related genes (Cancer Gene Census) in RIS genomes .....	24
<b>Figure 9.</b>	Enriched molecular functions of cancer-related genes harboring somatic mutations .....	26
<b>Figure 10.</b>	Genes undergone positive selections with respect to dNdSCV .....	27
<b>Figure 11.</b>	Genes harboring no less than two germline variants in RIS .....	29
<b>Figure 12.</b>	Enrichment of nine gene sets with DNA damage-repair pathways...	30

<b>Figure 13.</b> Somatic mutations of DDR genes in RIS genomes.....	31
<b>Figure 14.</b> Genome-wide somatic copy number alterations profiles of RIS genomes .....	34
<b>Figure 15.</b> Locus-level heatmaps of copy number alterations for GISTIC peaks .....	37
<b>Figure 16.</b> Frequencies of structural variations (duplication, deletion, inversion, and translocation) .....	40
<b>Figure 17.</b> Circos plots illustrating all structural variations of 11 RIS genomes	41
<b>Figure 18.</b> Microhomology ratio of structural variant breakpoints in 3 cohorts..	42
<b>Figure 19.</b> Example of chromothripsis event occurred in chr6 of RIS9.....	43
<b>Figure 20</b> Comparison of somatic mutations between RIS and sporadic sarcomas from TCGA.....	46



# List of Tables

<b>Table 1.</b>	Clinicopathological features of 11 RIS patients.....	9
<b>Table 2.</b>	Sequencing information of 11 RIS genomes.....	11
<b>Table 3.</b>	Enrichment analysis of DDR pathway with germline variants .....	32
<b>Table 4.</b>	GISTIC peaks of focal amplifications/deletions.....	35
<b>Table 5.</b>	SVs involving DDR genes in RIS genomes.....	44

# Introduction

Ionizing radiation is an established risk for developing cancers (1). Radiation induced cancers may appear after tens of years after radiotherapy in the form of cancers unrelated with the primary cancers. A linear dose-response relationship (2) indicates that the therapeutic radiation induced the secondary tumor, but the molecular mechanisms as to how ionizing radiation damages genomes and leads to the development of radiation induced sarcomas (RIS), are not well understood.

Radiation is a mutagen generating various types of DNA damages as demonstrated in experimental system. Various types of DNA damage can be induced by ionizing radiations (3). Thus, it is likely that the radiation hit has initially induced the genomic instability followed by generation of various types of genomic alterations, some of which involving key cancer-related genes acquiring driver roles. Frequent mutations in canonical tumor suppressor genes such as *TP53* and *RBI* (4) as well as transcriptional changes indicative of chronic stress have been previously reported in RIS (5).

Whole genome sequencing (WGS) presents the highest resolution cancer genome genotyping enabling the identification of small-to-large scaled genomic aberrations in the coding and noncoding regions of cancer genomes as recently demonstrated by a work of the international consortium (6). A recent study analyzed WGS data of 12 radiation-associated secondary malignancies including 9 RIS cases discovering the dominance of deletion over insertions along with an overrepresentation of balanced inversions (7). In addition, exome-scaled sequencing of RIS also revealed unique mutation signatures characterized by C>T transversions (8) suggesting that the high-

throughput sequencing data may reveal previously unrecognized features or biomarkers of RIS genomes.

The molecular characterization of the impact of ionizing radiation and the consequences have been largely limited because of the rare occurrence of RIS. Instead of using animal or cell line-based models, we have analyzed 11 sporadic cases of human RIS stringently defined. In this study, we performed WGS for ten RIS genomes with matched normal DNA. Four types of genomic alterations were identified including single nucleotide variations (SNV), short insertion/deletion (indel), somatic copy number alterations (SCNA), and structural variations (SV) including chromosomal translocations. SNVs and indels are collectively referred as somatic mutations in this study. Genes harboring various types of somatic alterations, including cancer-related genes, were identified. In addition, the abundance of somatic mutations along with the types of mutation signatures and SVs were investigated and compared with those of spontaneous osteosarcomas obtained from a public resource.

# Materials and Methods

## *Patient samples*

After the approval by the institutional review board of Seoul National University Hospital (IRB No: H-1506-026-678), patients were enrolled with informed consent. We identified patients who met the following criteria for RIS; (i) histologically different from primary cancer, (ii) developed in the irradiated field, and (iii) occurred with at least 6 months following radiotherapy (9). Between 2015 and 2019, eleven patients received resection. Tumor, adjacent normal tissues, formalin-fixed paraffin-embedded (FFPE), or blood were obtained from patients, and tissue blocks were stored in liquid nitrogen. Two researchers reviewed specimens, and microdissection was performed to isolate tumor and normal tissues. Clinical information including age, gender, latent period, treatment history, and date of RIS diagnosis was also collected.

## *DNA library preparation and whole-genome sequencing*

Genomic DNA was extracted from tumor and matched normal specimens using QIAamp DNA kit (Qiagen, Germany). NanoDrop 2000 (Thermo, USA) was used to determine the DNA concentration and quality with agarose gel electrophoresis. Sequencing library was prepared as previously described (10). Sheared genomic DNA 500ng was purified with end repair, 3- end adenylation, adaptors ligation, purification of ligation products. PCR amplified products were sequenced with Illumina X10 yielding paired-end sequencing of 151bp X 2. The sequencing reads were aligned onto the reference genome (hs37d5 used in the 1000 Genomes Project)

using BWA-MEM (v0.7.13) (11). Mapped reads were further processed for the local realignment and score recalibration using GATK (Genome Analysis ToolKit) (v3.5.0) pipeline (12). The processing of the sequencing data including the removal of duplicate and read sorting/indexing were performed using Samtools (v1.10) (11).

### ***Whole-exome sequencing***

Whole-exome sequencing was performed for the genomic DNA obtained from tumor and matched normal specimen using the Agilent SureSelect Human All Exome kit (Agilent Technologies, USA) and Illumina Novaseq 6000 platform. The depth of coverage of both tumor and normal was 200 X.

### ***Variant calling***

Somatic mutations were identified by comparing the tumor and matched normal WGS data. For variant calling, we used Sanger pipeline of International Cancer Genome Consortium (ICGC) proposed as standard ICGC-PCAWG (PanCancer Analysis of Whole Genome) mapping pipeline (6). Caveman algorithm (cancer variants through expectation maximization) (v1.5.0) was used to identify SNVs (13) exploiting the copy number profiles and related information including the tumor purity and ploidy estimated by ASCAT. Variants in known problematic regions or non-primary chromosomes were filtered out along with germline variants from 59 normal panels. Pindel (v1.5.7) algorithm was used to identify indels (14). To filter our common or mapping-related artifacts, the indels in normal panels and known problematic regions were also filtered out. ANNOVAR (2015Dec14) was used for variant annotation and functional prediction of somatic mutations (15). BRASS (BReakpoint via ASSEmbley) (v4.012) was used to identify simple and complex

genomic rearrangements with breakpoints. Fold-back artifacts and mismapping-associated false calls were filtered out as recommended (<https://github.com/cancerit/BRASS>). Copy number alterations were identified using ASCATngs (v1.5.2) using tumor-vs.-normal read depth ratios (16). Tumor purity and ploidy was also estimated and used for optimization of SNV and indel calling. The visualization of genome-wide SCNA profiles was performed using IGV browser (v2.8.4) (17). Focal, recurrent SCNAs of RIS genomes were identified using GISTIC 2.0 (18). SV visualization was done using circlize R package (v0.4.10) (19).

### ***Predisposing germline variants of RIS genomes***

We used HalpotypeCaller (GATK-v4.1.3.0) in GATK pipeline for normal genome WGS to identify the germline genetic variants in RIS patients. Germline variants with the allele frequencies > 1% in the public database including 1000 genomes project (<http://www.1000genomes.org>), ExAC (<http://exac.broadinstitute.org>), or NHLBI ESP exomes (<http://evs.gs.washington.edu/EVS/>) were filtered out. For 270 genes in nine DNA damage-repair (DDR) pathways (20), we identified 23 truncating germline variants (five nonsense, seven splicing sites and 11 frameshifting indels) and 50 missense variants called as potential damaging variants for no less than three times in five tests such as SIFT, PolyPhen2, MutationTaster, and MutationAssessor softwares. The enrichment test was done to estimate the significance of enrichment for seven DDR pathways in genes harboring truncating germline variants using Fisher's exact test.

### ***Chromothripsis***

We used ShatterSeek (21) to identify and visualize the *chromothripsis* as massive

genomic rearrangements involving oscillating copy number alterations for individual chromosomes. We used two parameters derived of ShatterSeek to discern *chromothripsis* for individual chromosomes, (i) the statistically significant enrichment of rearrangement breakpoints in given chromosomes ( $P < 0.05$ ) and (ii) the presence of oscillating copy number states for more than 4 segments.

### ***Mutation signatures***

*De novo* mutation signature discovery was done by applying non-negative matrix factorization (22) on the frequency matrix of trinucleotide context. Cophenetic score in permutation tests were used to estimate the optimum number of *de novo* mutation signatures. COSMIC mutation signatures of single nucleotide substitution (SBS) and indel (ID) mutation signatures were used for deconvolution of somatic mutations in RIS genomes (<https://cancer.sanger.ac.uk/cosmic/signatures>). For the deconvolution of SBS and ID mutations signatures, we used *deconstructSig* (23) and *sigminer* R packages (24), respectively. Mutation signature analysis was also performed for additional cohorts of RIS (9 RIS WGS data) (7), and 44 primary osteosarcoma (OSA) WGS data as available in ICGC consortium (6). For SBS and ID signatures, only top frequent signatures are shown while those with low frequencies were combined and collectively annotated as “others”.

### ***Comparison with Sporadic Sarcomas from TCGA***

Publicly available The Cancer Genome Atlas Program (TCGA) sarcoma (SARC) dataset was downloaded using the UCSC Cancer Genomics Browser (<https://genome-cancer.ucsc.edu>). We included patients who did not receive radiation with specific histologic types such as undifferentiated pleomorphic

sarcoma (n = 48) and myxofibrosarcoma (n = 25). The datasets were generated by Washington University using Illumina HiSeq 2000. We compared WGS data of TCGA and of ours with Maftools packages in R software.



# Results

## *Patients and data*

A total of 11 RIS patients who underwent curative resection were enrolled in the study. The clinicopathological information of the patients are shown in **Table 1**. We performed WGS for 11 pairs of tumor genome DNA and matched normal DNA of blood to identify somatic and germline genomic alterations of RIS genomes. The sequencing information including the sequencing depth is available in **Table 2**. Using the WGS, we identified four types of somatic alterations including single nucleotide variants (SNVs) and short insertions/deletions (indels), somatic copy number alterations (SCNA) and structural variations (SV) along with predisposing germline variants.

**Table 1. Clinicopathological features of 11 RIS patients**

	Sex	Age at primary RT (year)	Secondary tumor	Primary tumor	Latency (year)	Chemotherapy for primary	RIS location
RIS 1	F	46	Undifferentiated spindle cell sarcoma	Myxoid liposarcoma	4	No	Thigh
RIS 2	F	67	Undifferentiated pleomorphic sarcoma	Aggressive fibromatosis	7	No	Forearm
RIS 3	F	40	Undifferentiated pleomorphic sarcoma	Hodgkin lymphoma	21	Yes	Neck
RIS 4	M	58	Osteosarcoma	Rectal cancer	12	Yes	Pelvis
RIS 5	M	54	Undifferentiated spindle cell sarcoma	Hepatocellular carcinoma with bone metastasis	15	No	Humerus

RIS 6	F	66	Angiosarcoma	Cervix cancer	5	Yes	Thigh
RIS 7	F	54	Myxoid liposarcoma	Cervix cancer	7	Yes	Pelvis
RIS 8	F	75	Osteosarcoma	Leiomyosarcoma	10	No	Femur
RIS 9	F	27	Osteosarcoma	Cervix cancer	37	Yes	Pelvis
RIS 10	F	53	Undifferentiated pleomorphic sarcoma	Rectal cancer	9	Yes	Pelvis
RIS 11	M	60	Undifferentiated epitheloid sarcoma	Rectal cancer	16	Yes	Pelvis

---

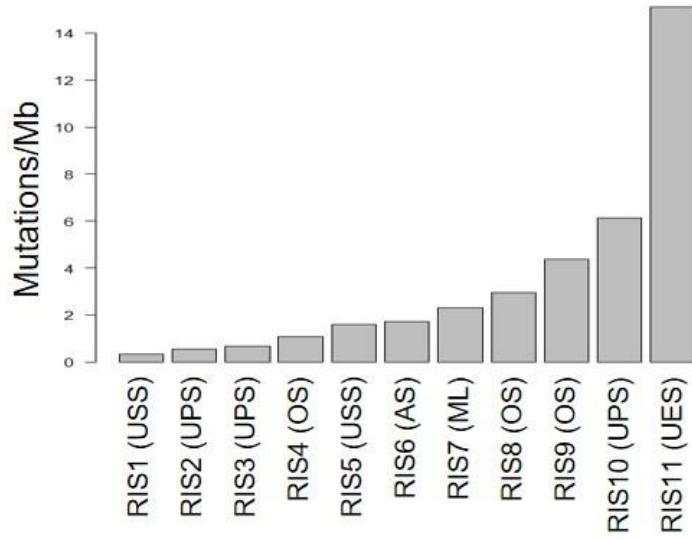
**Table 2. Sequencing information of 11 RIS genomes**

No	Initial Read	Processed Read	Mapped Read	Mapping Rate (%)	Average Depth	Read Length
RIS 1	2,314,920,174	2,189,106,265	2,187,454,798	99.92%	113.35	151
	1,396,578,196	1,328,136,369	1,325,888,066	99.83%	68.55	
RIS 2	2,185,981,260	2,015,891,886	2,007,768,647	99.60%	103.56	
	1,565,654,938	1,428,284,650	1,421,677,617	99.54%	73.27	
RIS 3	2,176,047,298	2,023,638,077	2,020,506,123	99.85%	104.53	
	1,701,026,272	1,627,360,166	1,623,622,170	99.77%	83.91	
RIS 4	2,105,933,820	1,818,469,275	1,808,891,044	99.47%	51.52	
	2,084,559,506	1,609,788,888	1,607,107,116	99.83%	73.29	
RIS 5	2,300,345,290	2,167,216,064	2,160,571,561	99.69%	111.21	
	1,445,722,764	1,368,539,317	1,366,647,491	99.86%	70.21	
RIS 6	3,129,314,730	2,397,831,736	2,393,780,638	99.83%	112.31	
	1,936,678,986	1,056,947,750	991,498,787	93.81%	42.24	
RIS 7	2,012,110,502	1,573,369,303	1,474,436,514	93.71%	56.08	
	1,260,735,898	1,199,145,292	1,195,386,718	99.69%	48.40	

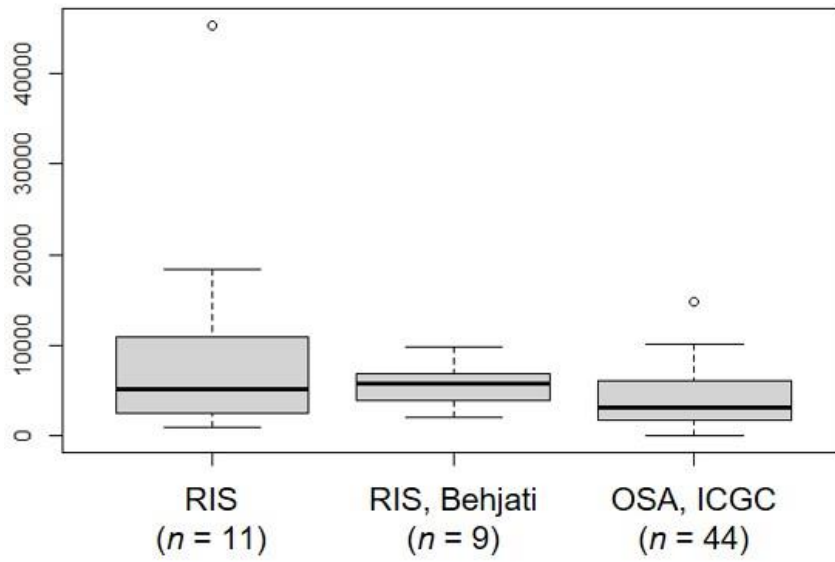
RIS 8	2,733,262,834	2,420,527,563	2,412,976,871	99.69%	124.66
	1,544,444,090	1,410,622,782	1,407,891,668	99.81%	72.95
RIS 9	2,246,782,100	1,898,661,719	1,894,074,200	99.76%	98.31
	1,687,965,900	1,538,364,566	1,533,818,154	99.70%	79.27
RIS 10	2,766,667,210	2,829,417,092	2,819,168,177	99.64%	92.79
	1,533,813,362	1,412,326,404	1,410,384,269	99.86%	70.29
RIS 11	1,732,807,728	1,625,837,382	1,621,674,095	99.74%	80.88
	1,280,231,274	1,209,592,017	1,200,858,763	99.28%	34.90

### ***The mutation abundance and signatures of RIS genomes***

We first investigated the abundance of somatic mutations in terms of tumor mutation burdens (TMB) in RIS genomes (**Figure 1**). RIS genomes are annotated in order of TMB (RIS1-RIS11). We identified one hypermutated RIS genome (RIS11 with 45,302 mutations corresponding to 15.1 mutations/Mb) while the other RIS genomes showed a wide range of TMB (RIS1 to RIS10; 956 – 18,427 mutations corresponding to 0.32 to 6.1 mutations/Mb). Given that the hypermutated genomes with high level of TMB are currently eligible to immune checkpoint inhibitors (25), our study suggests that some of RIS cases may be benefited from the immune checkpoint inhibitors. The TMB was also compared with those of 9 RIS in other cohort (RIS, Behjati) (7) and 44 spontaneously occurring osteosarcomas from the International Cancer Genome Consortium (OSA, ICGC), respectively (6). We observed that primary sarcomas showed reduced level of TMB compared to RIS genomes although statistically insignificant ( $P = 0.304$  and  $P = 0.085$  with RIS and RIS/Behjati, respectively,  $t$ -test). This finding suggests that the level of TMB may not be different between the radiation-induced from radiation-naïve sarcomas, but it requires further validation with extended cohorts (**Figure 2**).



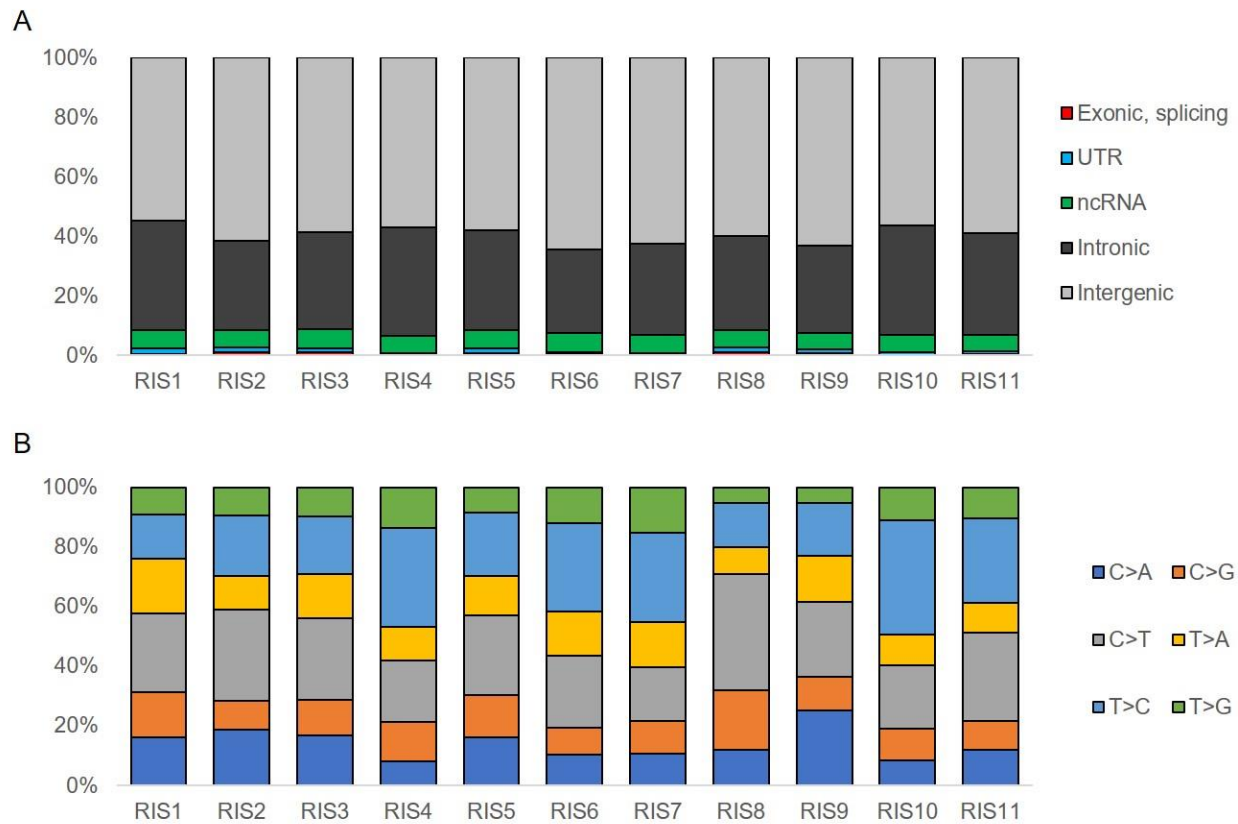
**Figure 1. Tumor mutation burdens (TMB) in RIS genomes**



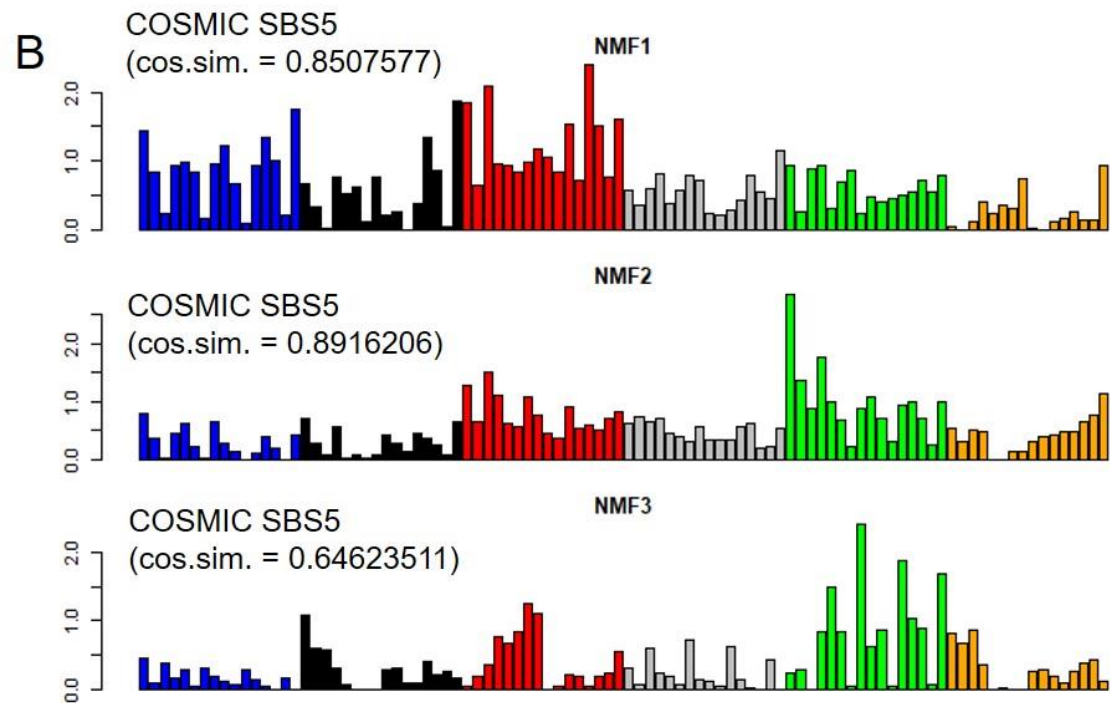
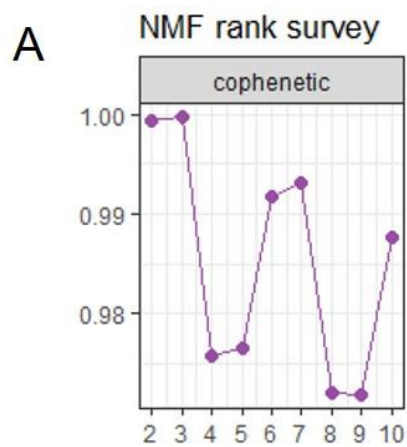
**Figure 2. Comparison of TMB between RIS and radiation- naïve sarcomas**



Next, we examined the abundance of mutations with respect to nearby genes and six mutation spectra (**Figure 3A** and **3B**, respectively). No substantial difference was observed between the RIS genomes for the intergenic-intronic mutation abundance. But variable extent of six mutation spectra was observed suggesting that the mutation context may be variable across RIS genomes. To further analyze the mutation spectra, we performed mutation signature analyses to deconvolute the somatic mutations in RIS genomes. Unsupervised deconvolution 96 trinucleotide context frequencies of SNVs based on non-negative matrix factorization (NMF) revealed there might be three SNV mutation signatures (**Figure 4**). The three identified mutation signatures were similar with each other and highly correlated with known mutation signatures of SBS5, clock-like signatures with unknown causal relationships (COSMIC single base substitutions or SBS). This incomplete deconvolution may be due to the small number of cases.

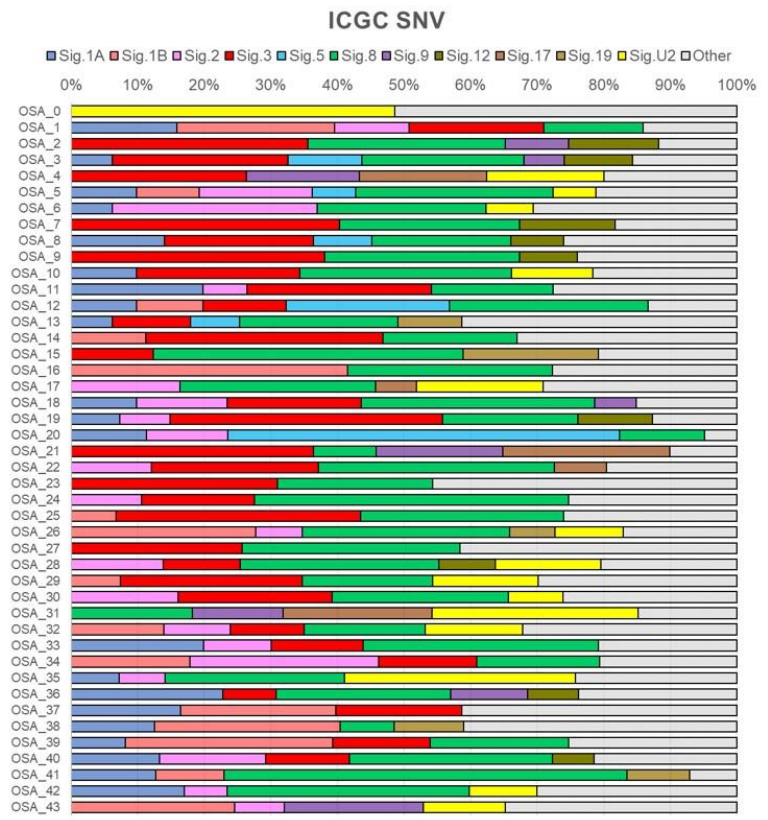
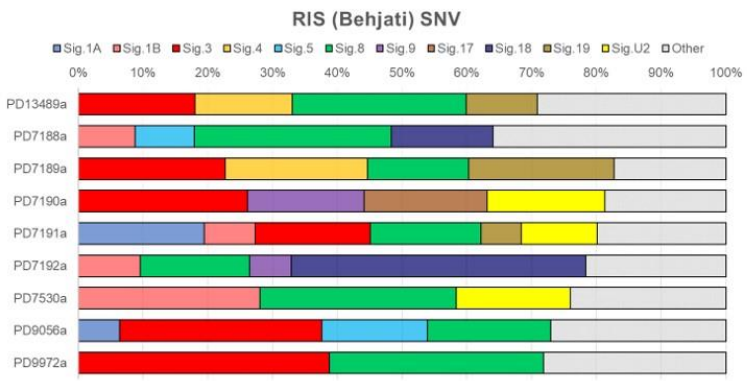
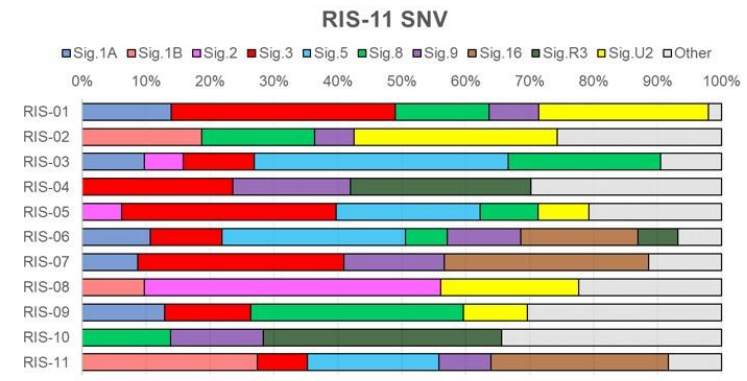


**Figure 3. Mutation features with respect to nearby (A) genes and (B) spectra in 11 RIS genomes**

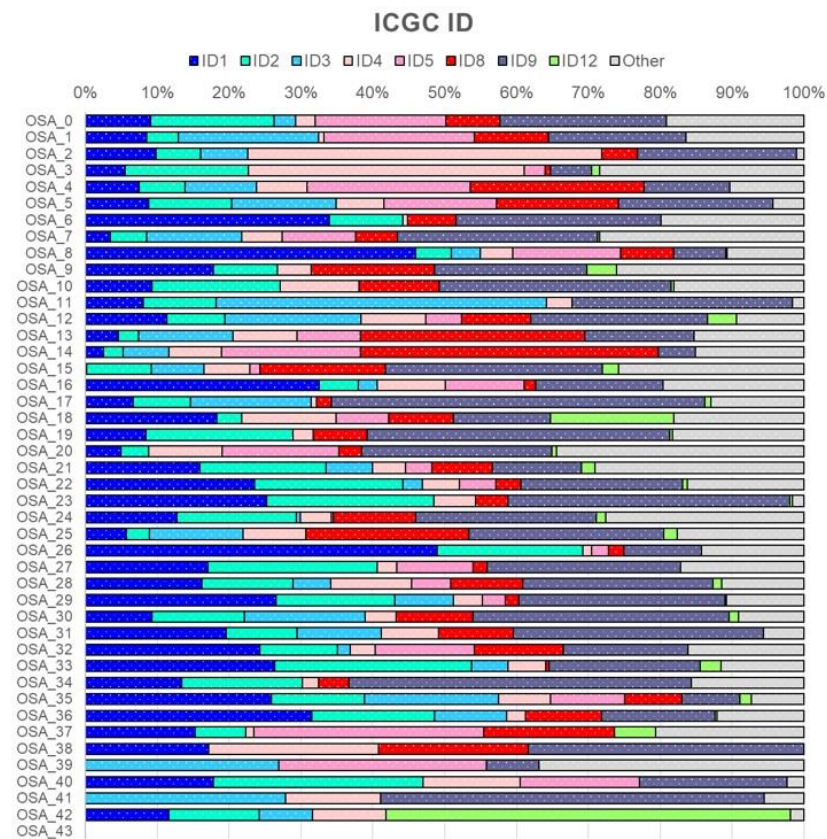
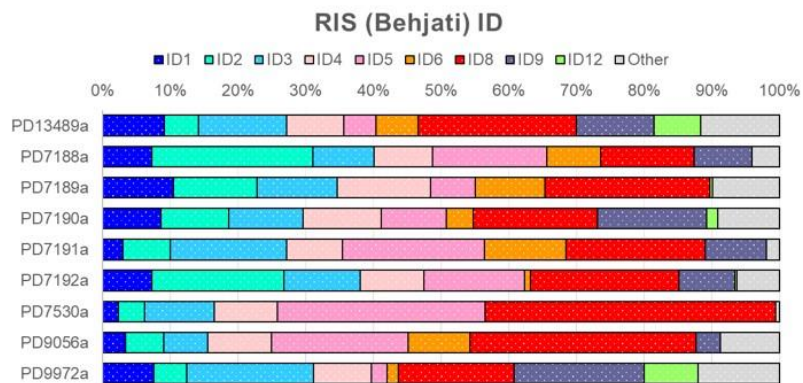
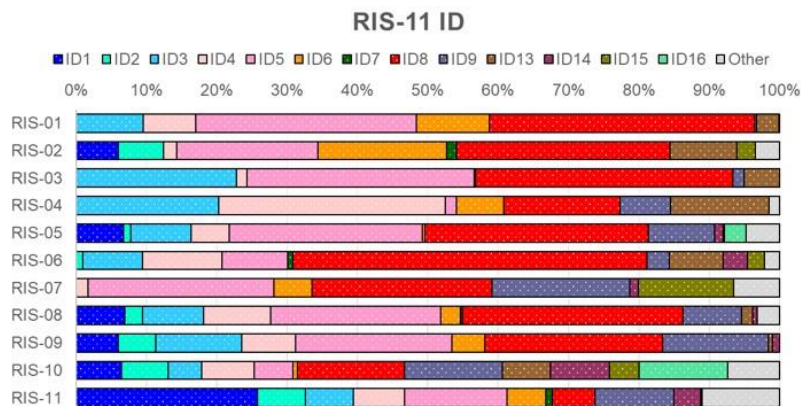


**Figure 4. De novo discovery of mutation signatures in 11 RIS genomes**

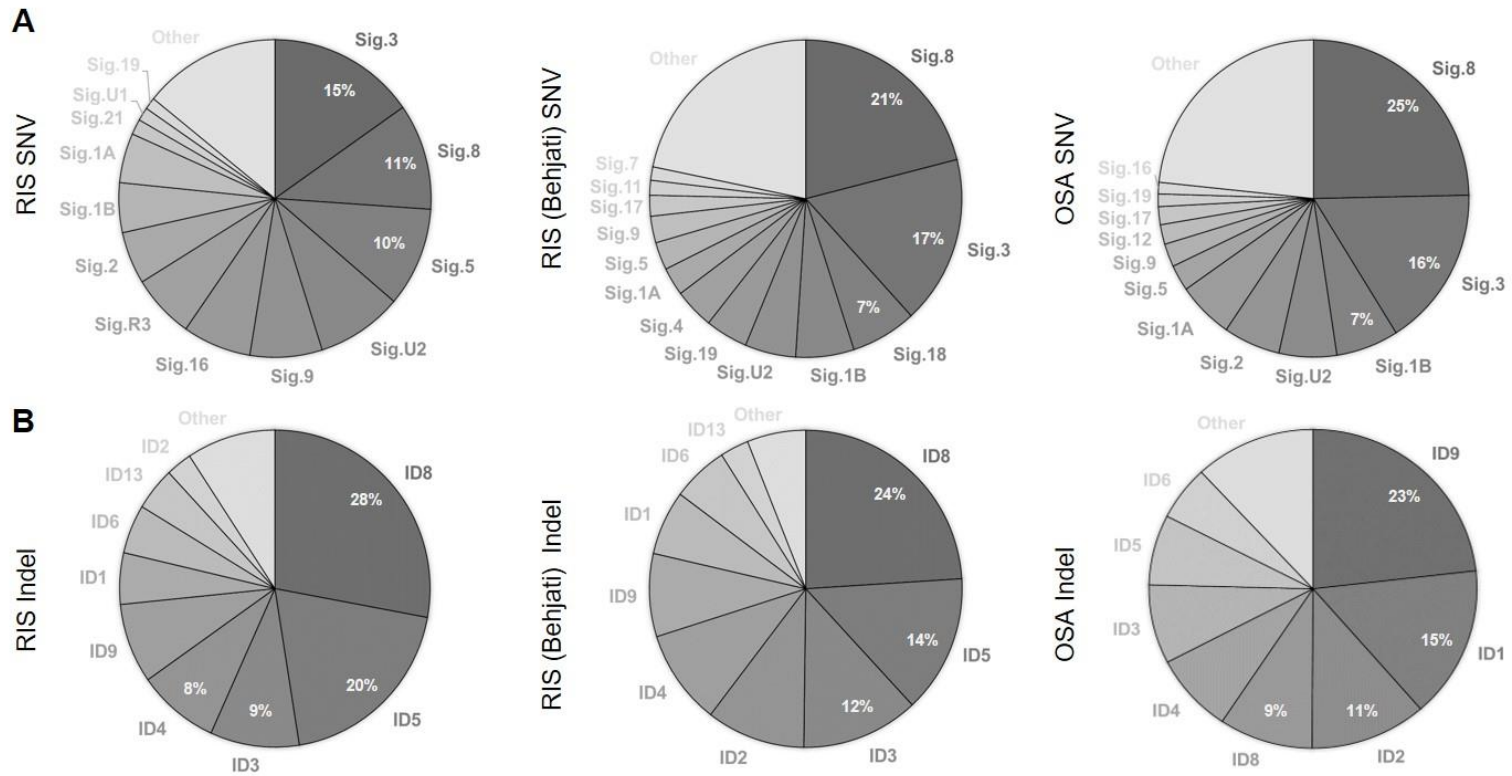
Therefore, we next performed supervised mutation signature analysis. SBS and ID (indel) COSMIC signatures (**Figure 5** and **6**, respectively) and compared the results with those of other RIS and OSA cohorts (**Figure 7**). For SBS signatures, we observed that RIS and OSA genomes showed similar signature profiles, e.g., SBS1, SBS3 and SBS8 are consistently over-represented both in radiation-induced and -naïve sarcomas. SBS1 and SBS3 have been well-recognized for their causalities, e.g., SBS1 mutations as products of 5'-methylcytosine deamination are associated with patients' age (26) and SBS3 mutations arise in the context of deficient of homologous recombination often with *BRCA* mutations (27). SBS8 signatures represent the excess accumulation of 8-oxo-2'-deoxyguanosine as well-recognized oxidative base lesions arising with ionizing radiations (28), but we also observed that this signature is also enriched in OSA cohort suggesting SBS1, SBS3 and SBS8 mutations may arise independent of radiation. Although less frequent, SBS5 mutations representing frequent A·T → G·C transitions were relatively enriched in RIS genomes suggesting that RIS genomes may be subject to excessive deamination such as adenine to xanthine (29). In addition, RIS (Behjati) genomes were specifically enriched with SBS18 that has been associated with oxidative DNA damage due to reactive oxygen species and also commonly observed in mouse tumors with high-energy radiation (28). Of note, RIS genomes commonly showed the enrichment of ID signatures of ID8, ID5 and ID3 while OSA genomes were enriched with ID1, ID2 and ID9 signatures (**Figure 7**). This notable difference of ID signature abundance between RIS and OSA genomes suggests that the mechanisms giving raise to indels may be differential between two types of cancers. For example, top enriched ID signatures of RIS genomes (ID8 signature) is associated with DNA double strand breakage repair by non-homologous end-joining (NHEJ), whereas ID1 and ID2 signatures of OSA genomes are associated with DNA mismatch repair deficiency.



**Figure 5. Mutational signatures of SNV in 3 cohorts**



**Figure 6. Mutational signatures of indel in 3 cohorts**



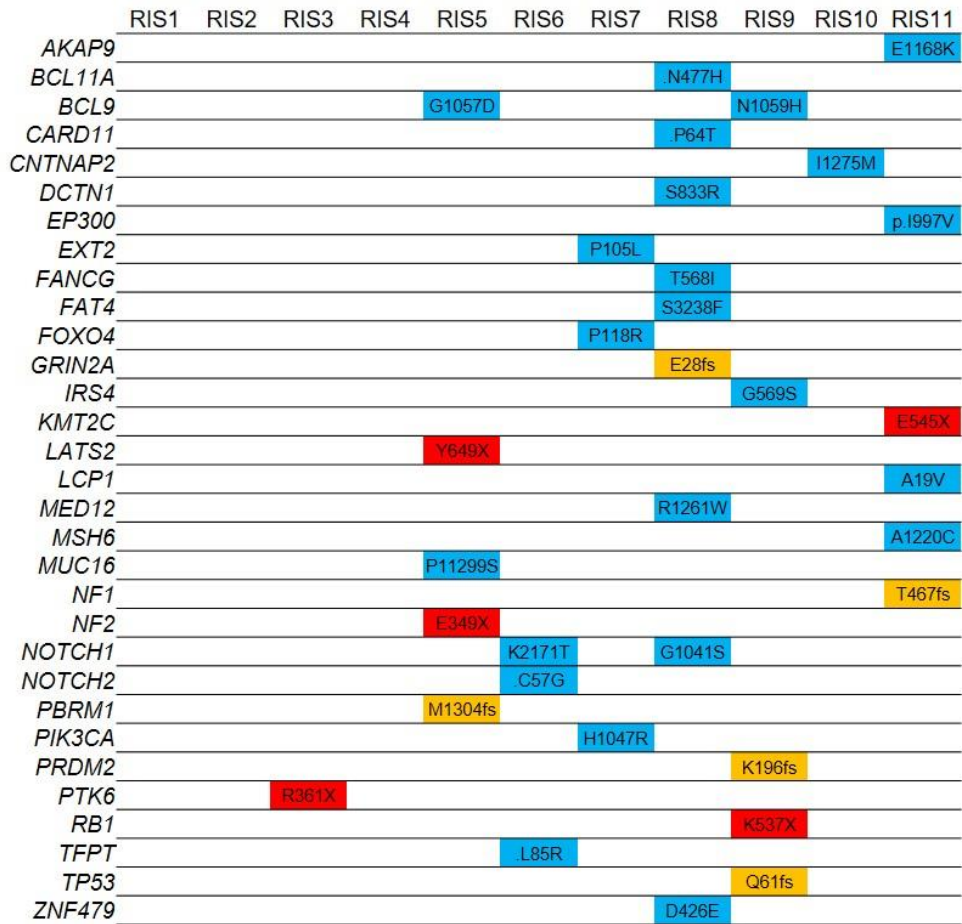
**Figure 7. Comparison of COSMIC (A) SBS and (B) ID mutation signatures between RIS and radiation- naïve sarcomas**



### ***Functional mutations and potential drivers in RIS genomes***

We next examined the potential driver mutations as those of cancer-related genes in RIS genomes. **Figure 8** shows the mutation landscape of RIS genomes for those observed in cancer-related genes (Cancer Gene Census) (30). Recurrent mutations on cancer-related genes of *BCL9* and *NOTCH1* were observed. The deregulation of *BCL9*, as component of Wnt signaling, may have oncogenic roles in RIS development (31). Although two *BCL9* missense mutations occurred outside of CTNNB1 binding domains (amino acids residues 358-374), their physical proximity (p.G1057D and p.N1059H) is suggestive of potential functionality. *NOTCH1* mutations are widespread across the cancers with dual roles as protooncogenes and tumor suppressors (32). Two observed missense *NOTCH1* mutations (G1041S in EGF-like domain and K2171T in C-terminal of the encoded peptide) may represent activating events whose transcriptional up-regulation has been reported in soft tissue sarcoma (33).

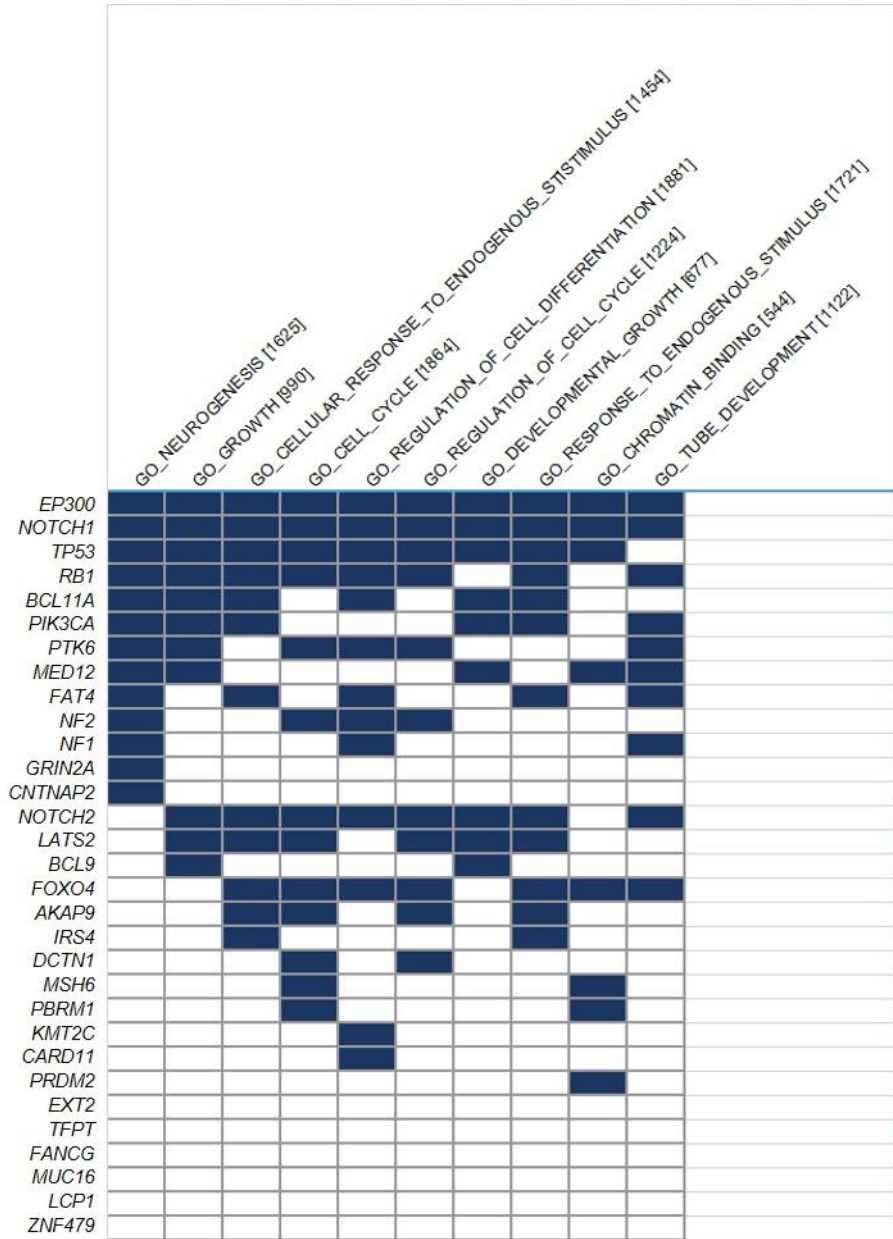




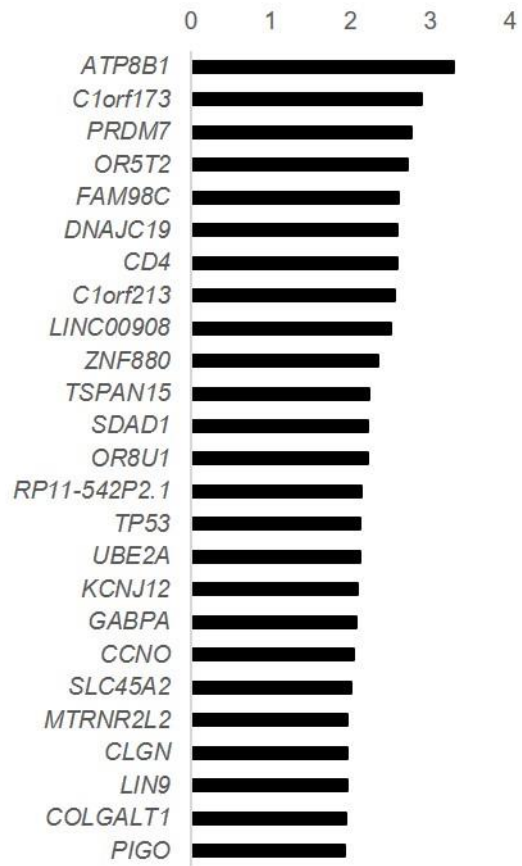
**Figure 8. Somatic mutations observed in cancer-related genes (Cancer Gene Census) in RIS genomes**

In addition, singleton mutations in genes with relatively well-recognized roles were observed in RIS genomes. For example, loss-of-functional, truncating mutations were observed on *NF2*, *RBI* and *TP53* mutations with potential roles in the RIS development along with a known hotspot missense mutation of *PIK3CA* (p.H1047R) (34). Truncating mutations of *LATS2* were observed and may be involved in RIS carcinogenesis as component of Hippo pathway (35). When the cancer-related genes harboring somatic mutations were examined for the enrichment to Gene Ontology terms, the functions representing growth, cell cycle and chromatin binding were frequently associated (**Figure 9**).

For all genes harboring somatic mutations, we employed dNdSCV, a measure of positive selection of somatic mutations to identify potential RIS drivers genes (36). The top significant 20 genes with positive selections are listed in **Figure 10**. Among the genes identified, *PRDM7* with one missense and one nonsense mutations, encodes an epigenetic regulator of histone 3 lysine 4 trimethyltransferase (37). Although their roles in RIS development is not well recognized, frequent mutations of *PRDM7* along with other epigenetic regulators such as *KMT2C/MLL3*, *PBRM1* and *PRDM2* may be implicated as functional events in RIS development (38).



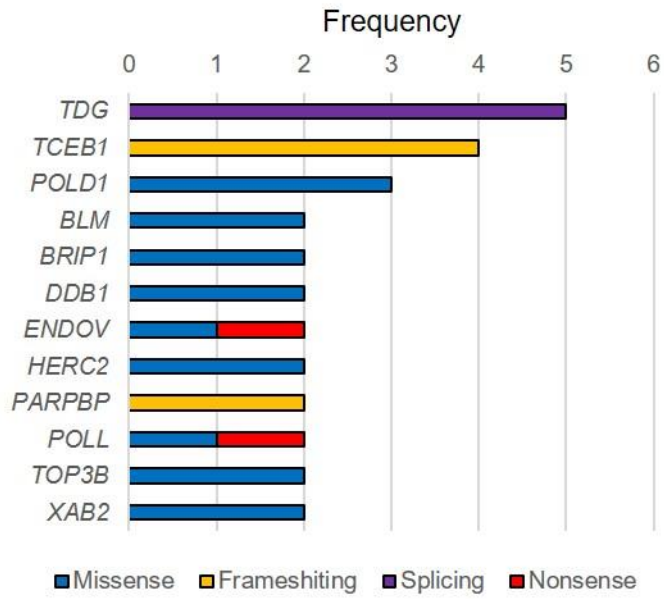
**Figure 9. Enriched molecular functions of cancer-related genes harboring somatic mutations**



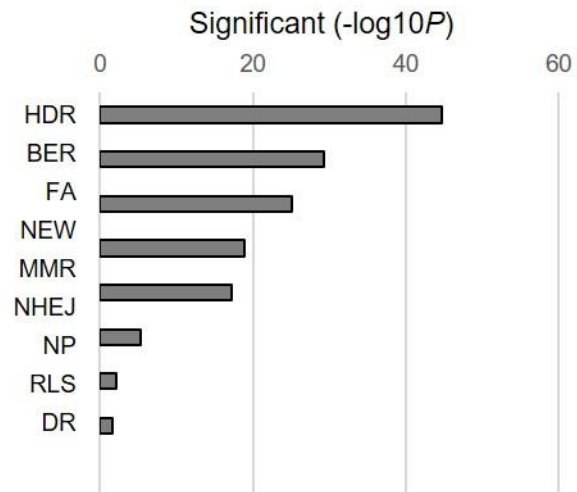
**Figure 10. Genes undergone positive selections with respect to dNdSCV**

### ***Predisposing germline variants of RIS patients***

We next investigated the germline variants that can be predisposing the development of RIS genomes. We focused on 270 in nine DDR pathways as likely candidates associated with the genomic instability and increased susceptibility to RIS (20). A total of 73 germline truncating or potentially damaging missense variants were observed 55 gene belonging to the DDR pathway. Eleven RIS genomes harbored 2 to 16 germline variants with the most frequent variants observed for *TDG*, *TCEBI* and *POLD1* (**Figure 11**). Genes harboring no less than two germline variants are shown in **Figure 11**. *TDG* encoding thymine DNA glycosylase that corrects deamination-induced DNA mismatches as a base excision repair enzyme harbored five splice-site mutations. For *POLD1* variants, one out of three missense germline variants were observed for the hypermutated case (RIS11) suggesting that the hypermutated genotype of RIS11 may be attributed with germline origins of *POLD1* deficiency (39). Then, we examined the significance of enrichment of genes harboring predisposing germline variants with nine DDR pathways (**Figure 12, Table 3**). Homology-dependent recombination was the most enriched DDR pathway ( $P = 2.4E-45$ ; Fisher's exact test) followed by base excision repair and Fanconi anemia ( $P = 4.7E-30$  and  $P = 0.6E-26$ , respectively). This suggests that the germline deficiency of these DDR pathway may be involved in RIS pathogenesis. Seven DDR genes also harbored somatic mutations (i.e., *FANCG*, *MSH6*, *PARG*, *RAD54B*, *TP53*, *TP53BP1*, and *UBE2A*) (**Figure 13**). We did not observe the evidence of biallelic losses accompanying germline variants for these somatic mutations, but it will require further evaluation for additional mechanisms of inactivation such as DNA promoter methylation or the haploinsufficiency. For example, *POLD1* germline mutation without 'second hits' would increase the mutation frequency because half of polymerase function is error-prone (40).



**Figure 11. Genes harboring no less than two germline variants in RIS**



**Figure 12. Enrichment of nine gene sets with DNA damage-repair pathways**



**Figure 13. Somatic mutations of DDR genes in RIS genomes**



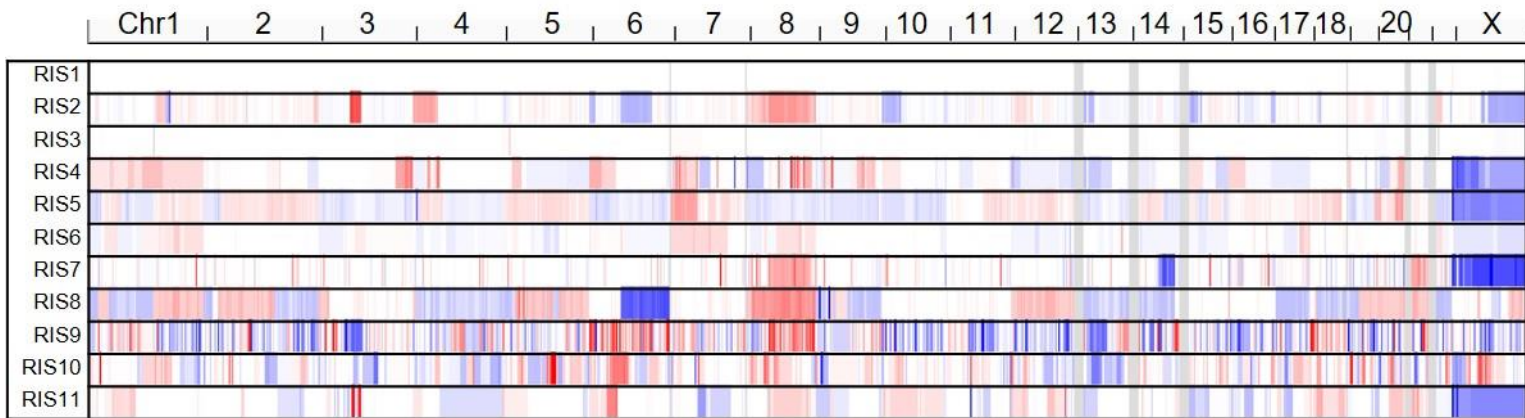
**Table 3. Enrichment analysis of DDR pathway with germline variants.**

DDR pathway	No. Genes	Set size	P values	Genes
HDR	22	88	2.41E-45	<i>BLM, BRIP1, EXO1, HELQ, LIG1, NSMCE4A, PALB2, PARPBP, POLD1, RAD51, RAD51B, RAD52, RAD54B, RECQL5, RFC1, RMI2, RPA1, RPA4, RTEL1, TOP3A, TOP3B, TP53BP1</i>
BER	14	47	4.69E-30	<i>APEX2, APTX, LIG1, LIG3, MUTYH, NEIL2, OGG1, PNKP, POLD1, POLE2, POLL, RFC1, TDG, XRCC1</i>
FA	12	41	9.63E-26	<i>BLM, BRIP1, FAAP100, FAAP20, FANCB, FANCC, HELQ, PALB2, RAD51, RMI2, TOP3A, TOP3B</i>
NER	10	51	1.24E-19	<i>DDB1, ERCC2, LIG1, POLD1, POLE2, RFC1, RPA1, RPA4, TCEB1, XAB2</i>
MMR	8	24	5.42E-18	<i>EXO1, LIG1, MLH1, PMS2, POLD1, RFC1, RPA1, RPA4</i>
NHEJ	3	23	4.27E-06	<i>PNKP, POLL, TP53BP1</i>
NP	1	5	6.86E-03	<i>NUDT15</i>
TLS	1	20	2.72E-02	<i>POLN</i>
DR	0	4	1	

Homology-dependent recombination (HDR), Base Excision Repair (BER), Fanconi Anemia (FA), Nucleotide Excision Repair (NER), Mismatch Repair (MMR), Non-homologous End Joining (NHEJ), Nucleotide pools (NP), Translesion Synthesis (TLS), Direct Repair (DR)

### *Somatic copy number alterations*

To identify large scale, chromosomal copy number changes in RIS genomes, we employed read depth-based SCNA calling algorithm (16). We observed that SCNA are prevalent in RIS genomes except for two genomes (RIS1 and RIS3) that were also devoid of somatic mutations. The genome-wide SCNA profiles of 11 RIS genomes are illustrated in **Figure 14** highlighting recurrent chromosomal SCNAs such as gains of 8q including *MYC* are prevalent across RIS genomes. To further identify the focal, recurrent SCNAs in RIS genomes, we used GISTIC algorithm (18) (**Figure 15**). A total of 9 GISTIC peaks were identified, a majority of which were focal deletion peaks ( $n = 8$ ) indicating that focal chromosomal deletions are frequent and likely to be recurrent in RIS genomes than focal amplifications. Focal deletions at 9p21.3 and 13q14.11 and 17q12 include canonical tumor suppressor genes of *CDKN2A*, *RBI* and *NFI*. For *RBI* and *NFI*, somatic mutations were also observed in our cohort. The GISTIC peaks are available in **Table 4**. Locus-level heatmaps of copy number alterations are shown for nine GISTIC peaks as well as for *MYC* locus in **Figure 15**. Comparing *RBI* and *NFI* deletions occurring in a limited number of cases, but *MYC* amplification is widespread across the cases.



**Figure 14. Genome-wide somatic copy number alterations profiles of RIS genomes**

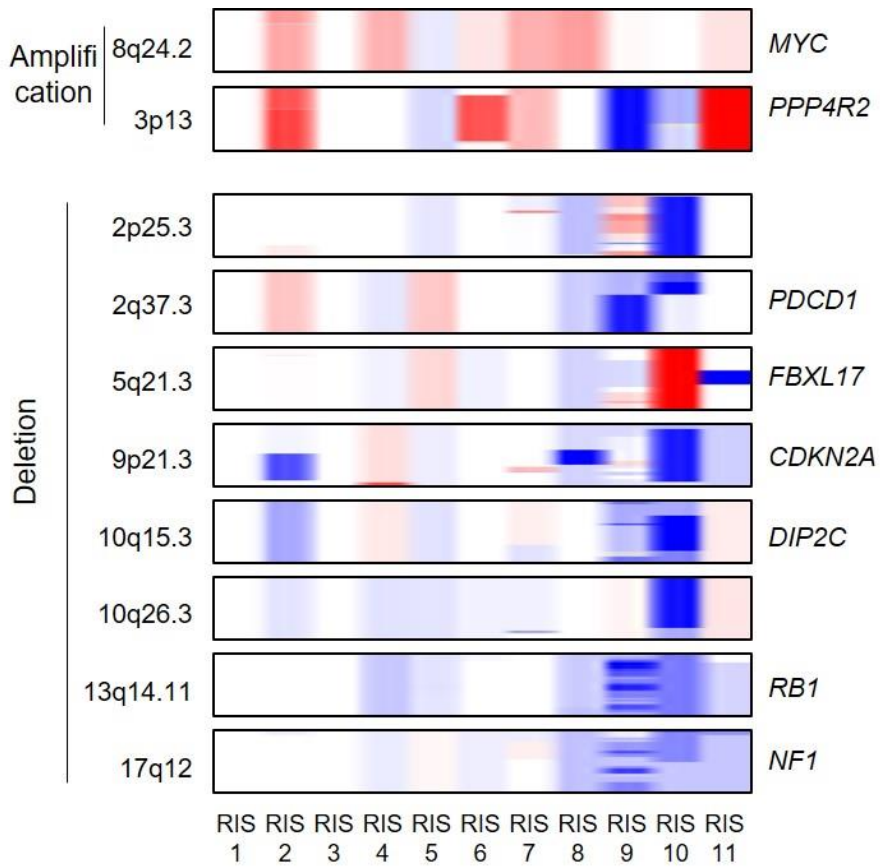
**Table 4. GISTIC peaks of focal amplifications/deletions**

Unique Name	Descriptor	Wide Peak Limits	Peak Limits	Region Limits	q values
Amplification Peak 1	3p13	chr3:73086586-73175404 (probes 59911:59922)	chr3:73095045-73145134 (probes 59912:59918)	chr3:69608837-75420642 (probes 59542:60174)	0.055228
Deletion Peak 1	2p25.3	chr2:1-120993620 (probes 26386:38975)	chr2:29444-875379 (probes 26389:26487)	chr2:1-937583 (probes 26386:26495)	0.035228
Deletion Peak 2	2q37.3	chr2:242695977-243199373 (probes 51929:51946)	chr2:242705903-243199373 (probes 51930:51946)	chr2:242725753-242935544 (probes 51932:51933)	0.14803
Deletion Peak 3	5q21.3	chr5:107004050-108092362 (probes 104099:104213)	chr5:107176733-107736691 (probes 104119:104175)	chr5:107426901-107736691 (probes 104147:104175)	0.086733

---

Deletion Peak 4	9p21.3	chr9:1-141213431 (probes 163153:178034)	chr9:21442379-22130514 (probes 165649:165730)	chr9:21475903-21499298 (probes 165654:165655)	0.24946
Deletion Peak 5	10p15.3	chr10:1-697036 (probes 178035:178095)	chr10:1-313424 (probes 178035:178052)	chr10:1-657135 (probes 178035:178091)	0.14803
Deletion Peak 6	10q26.3	chr10:135281745-135534747 (probes 192321:192337)	chr10:135326542-135534747 (probes 192327:192337)	chr10:135285338-135534747 (probes 192322:192337)	0.035228
Deletion Peak 7	13q14.11	chr13:34539866-52350616 (probes 223038:224965)	chr13:41282180-41721049 (probes 223762:223804)	chr13:34987019-50643298 (probes 223084:224783)	0.027738
Deletion Peak 8	17q12	chr17:26633322-44851047 (probes 262434:264346)	chr17:31840498-32980095 (probes 262986:263114)	chr17:31367270-33081000 (probes 262929:263126)	0.14803

---



**Figure 15. Locus-level heatmaps of copy number alterations for GISTIC peaks**

### ***Structural variations***

A total of 1,735 SVs were identified with their breakpoint ends in 11 RIS genomes (1 to 966 SVs per case with median of 42 SVs). Frequencies of four SV categories (duplication, deletion, inversion, and translocation) are shown in **Figure 16**. The RIS genomes relatively devoid of somatic mutations and SCNAs (RIS1 and RIS3) harbored frequent SVs. For majority of cases (8 out of 11 cases), the frequency of deletions dominates those of duplication, which has been proposed as RIS-specific genomic signatures (7). Circos plots are shown illustrating all the SV of given genomes (**Figure 17**). Variable abundance of SVs across the cases were observed ranging from a SV-depleted case (RIS7) to those with massive genomic rearrangements (RIS8, RIS9 and RIS11).

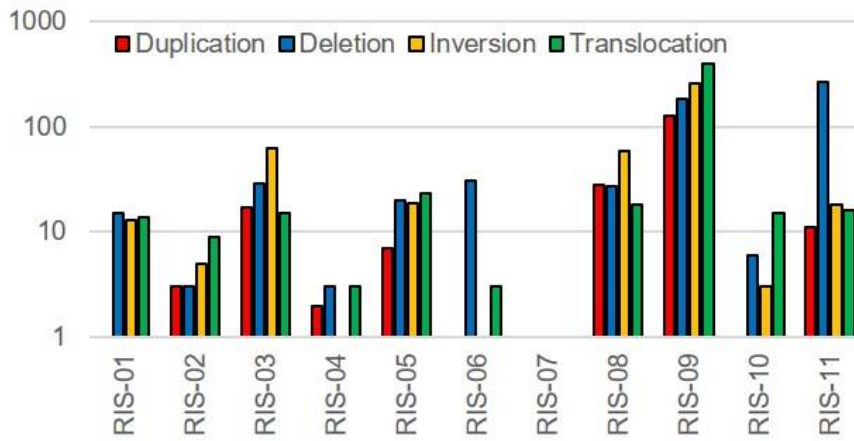
Cancer-related genes recurrently affects by SV in more than one case are *CNTNAP2*, *NF2*, *CTNND2*, *PTPRD*, *ARID1B*, *RAD51B*, *FOXP*, and *ZEB1*. Thus, *NF1* and *NF2* are frequently inactivated with different types of genomic alterations (SCNAs and SVs, respectively) in RIS genomes. Among epigenetic modulators, *ARID1B* may represent frequent targets of SVs (RIS6 and RIS9). Other protooncogenes associated with SVs are observed such as *CTNND2* requires further investigation as recently reported for their roles in tumor development (41). It is also noted that DDR gene of *RAD51B*, involved in homologous recombination harbors recurrent SVs (RIS9 and RIS11). *RAD51B* rearrangements have been previously observed in mesenchymal origin tumors (42) and given the potential predisposing roles in tumor development (43), it is possible that the *RAD51B* rearrangements may present the early genomic changes contributing to the RIS development. SVs involving 14 DDR genes including *RAD51B* are listed in **Table 5**.

We further identified the present of microhomology or non-template sequences at

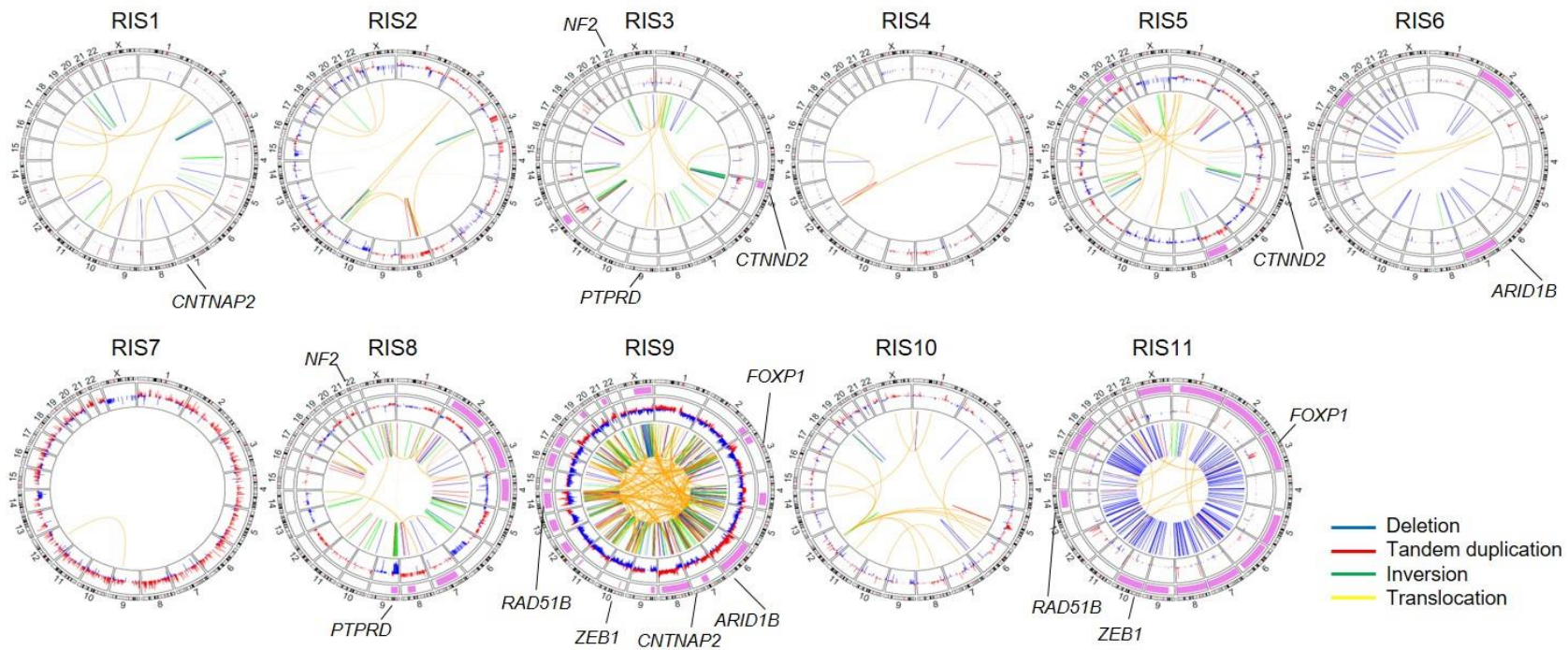
SV breakpoints. Across different SV types, RIS genomes frequently harbor microhomology or non-template sequences compared to OSA genomes as potential genomic hallmark of RIS (**Figure 18**). The prevalence of microhomology and non-template sequences at breakpoints suggests that the SV formation and repairs are driven by alternative end joining rather than homologous recombination or classical NHEJ (44).

In addition, *chromothripsis* as massive chromosomal rearrangements were observed more than half of the cases (RIS2, RIS3, RIS5, RIS6, RIS8, RIS9, and RIS11), consistent with prevailing chromosomal rearrangements in sarcomas (45). One example of *chromothripsis* (chr6 of RIS9) involving massive intrachromosomal rearrangements is illustrated in **Figure 19**.

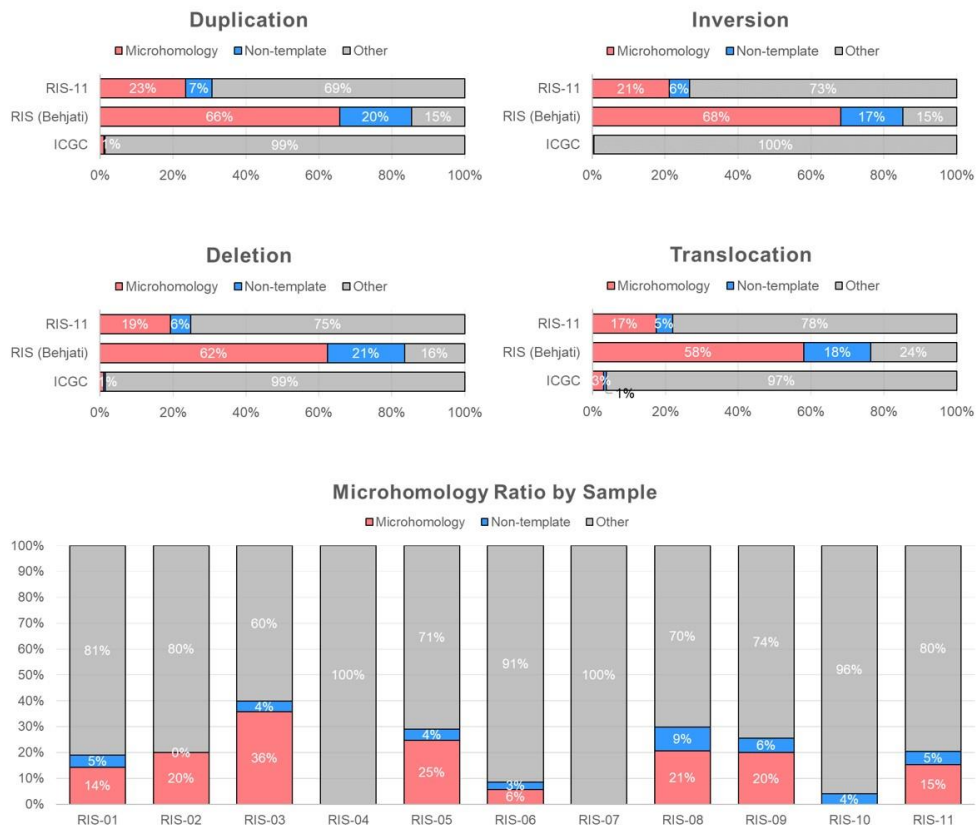




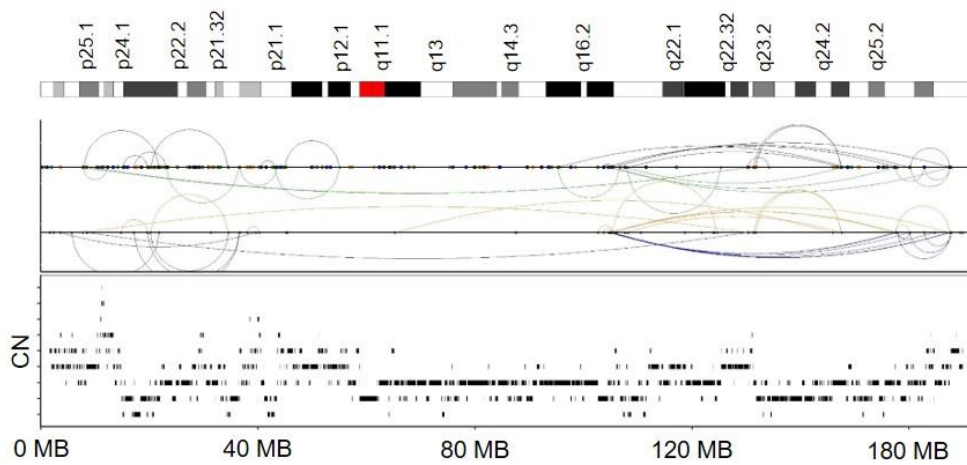
**Figure 16. Frequencies of structural variations (duplication, deletion, inversion, and translocation)**



**Figure 17. Circos plots illustrating all structural variations of 11 RIS genomes**



**Figure 18. Microhomology ratio of structural variant breakpoints in 3 cohorts**



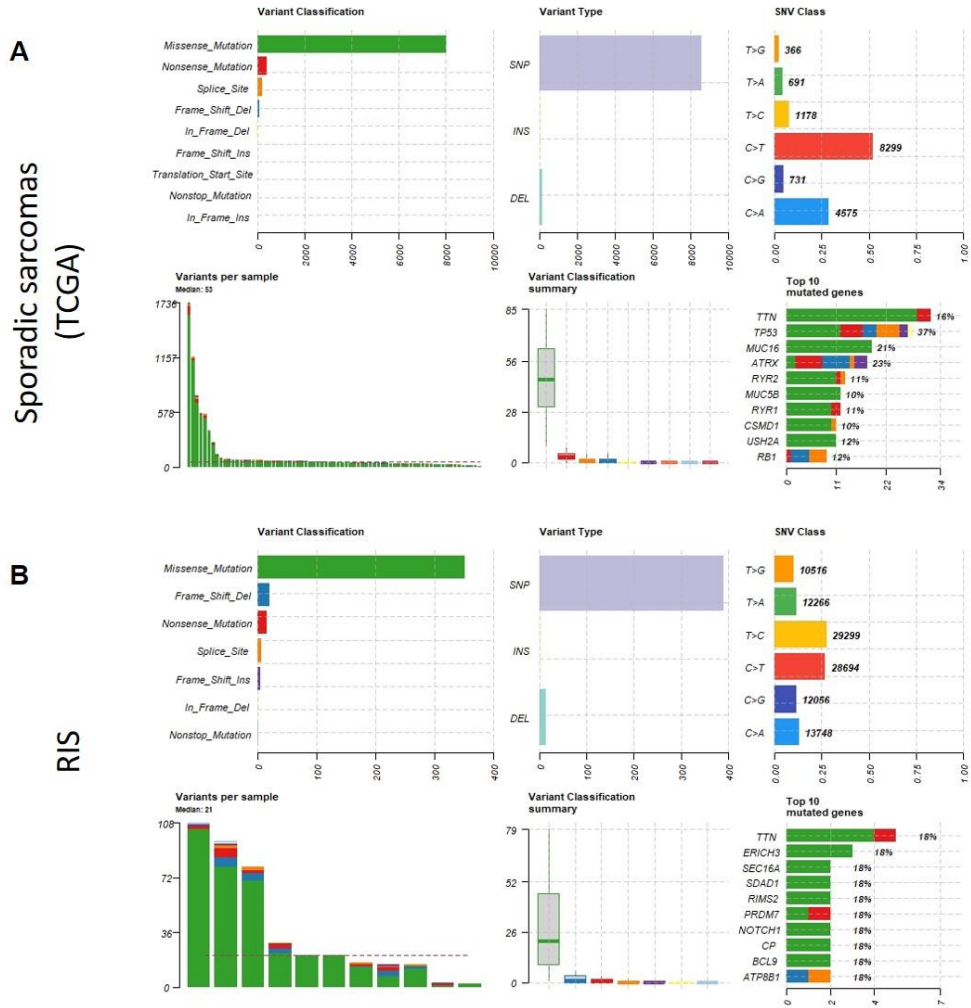
**Figure 19.** Example of *chromothripsis* occurred in chr6 of RIS9

**Table 5. SVs involving DDR genes in RIS genomes**

Sample	Gene	Group	SV Type
RIS-09	<i>ALKBH1</i>	Base Excision Repair (BER)	INV, TRA
RIS-09	<i>BRCA2</i>	Homology-dependent recombination (HDR)	DUP, TRA
RIS-09	<i>CUL5</i>	Nucleotide Excision Repair (NER)	INV
RIS-08	<i>FANCG</i>	Fanconi Anemia (FA)	INV
RIS-09	<i>MNAT1</i>	Nucleotide Excision Repair (NER)	TRA
RIS-09	<i>MSH3</i>	Mismatch Repair (MMR)	TRA
RIS-09	<i>PARP4</i>	Base Excision Repair (BER)	INV
RIS-09	<i>POLB</i>	Non-homologous End Joining (NHEJ)	TRA
RIS-09	<i>POLE2</i>	Base Excision Repair (BER)	TRA
RIS-09	<i>RAD51B</i>	Homology-dependent recombination (HDR)	TRA, INV, DUP, DEL
RIS-11	<i>RAD51B</i>	Homology-dependent recombination (HDR)	TRA, INV, DUP, DEL
RIS-11	<i>RAD52</i>	Homology-dependent recombination (HDR)	DEL
RIS-09	<i>RFC3</i>	Base Excision Repair (BER)	DUP, INV
RIS-09	<i>WDR48</i>	Fanconi Anemia (FA)	TRA

### ***Comparison with Sporadic Sarcomas from TCGA***

When comparing the sporadic sarcoma and RIS regards to somatic mutations, more frame-shift-deletion was observed in the RIS group than in the nonsense mutation compared to TCGA (**Figure 20**). Also, the number of variants was less in RIS than in TCGA group. Median number of variants per sample was 53 in sporadic sarcomas from TCGA and 21 in RIS. In the case of mutated gene, *TP53* was observed in 37% and *RBI* 12% of sporadic sarcomas, but not in RIS.



**Figure 20. Comparison of somatic mutations between RIS and sporadic sarcomas from TCGA**

## Discussion

It has been previously proposed that somatic mutations are relatively depleted compared to SCNAs or SVs in RIS, but the mutation burdens in terms of WGS scale have not been compared with spontaneous sarcomas. In this study, we demonstrated that the mutation burdens of RIS genomes were largely comparable to those of spontaneous sarcomas. Moreover, the frequency of pathogenic mutations in the cancer driver genes is relatively low in RIS genomes, suggesting that they may rise in an atypical pathway of tumorigenesis. These observations suggest that the mutation burdens of sarcomas may not be different with respect to the tumor causalities and also, somatic mutations may not be a major contribution to the RIS development. But, due to the known heterogeneities of sarcomas and the relatively small size of study cohorts, further investigation will be required.

Mutation signature analysis has recently shown that mutation profiles of given tumors can be deconvoluted into a set of relative contribution of different mutation processes (26). Our study shows that SBS and ID mutation signatures may provide clues on the mutational processes associated with RIS development and especially those of radiation-specific mutagenic impact. COSMIC SBS3 was prevalent in RIS genomes. It corresponds to a deficiency in homologous recombination and shows a good therapeutic response to poly(ADP-ribose) polymerase (PARP) inhibitors (27,46). The in-vitro study demonstrated that 55% of sporadic sarcoma samples carried *BRCA1/2* mutations and were also sensitive to PARP inhibitors (47). Tumors showing SBS3 without mutation in *BRCA1/2* or other significant HR-related genes may benefit from these selective inhibitors (48). Along with previous studies, our data suggest that PARP inhibitors could be a promising therapeutic candidate for RIS



showing SBS3.

A recent study of mouse radiation-associated secondary malignancies suggests that SBS18 representing oxidative stress may be associated with radiation-induced DNA damages especially for those with high level of radiation energy (28). In our study, SBS18 is observed in one of two RIS cohorts, suggestive of a substantial heterogeneity of RIS genomes. In addition, SBS8 that is also characterized by C → A transition with SBS18, was prevalent in RIS genomes suggesting that the ionizing radiation can damage the genomes indirectly via free radical formation leaving signatures of C → A transition. However, SBS8 was also prevalent in OSA genomes so that SBS8 mutations may arise from other mutagenic resources. Other mutation signatures enriched in RIS genomes include SBS5 characterized by excess of T → C transition. It will require further investigation since these signatures such as SBS5 and SBS8 are not well recognized for their causality, and further to validate whether the excess of C → A and T → C transitions may represent the genomic footprints of ionizing radiations and RIS genomes.

Notably, indel-associated ID signatures of RIS genomes were distinctive of OSA genomes proposing potential mechanisms giving rise to indels as well as genomic hallmarks associated with ionizing radiation. For example, ID8, characterized by deletions larger than 5-bp was commonly observed in RIS genomes, not in sporadic sarcomas. This deletion pattern has been reported in genomes subject to NHEJ repair for DNA double strand breaks induced by radiation (49). We can suggest that this indel signatures confirmed with RIS samples using WGS seems to be helpful in discriminating radiation-induced secondary malignancies from sporadic cancers.

Identifying high risk patients for secondary malignancy may benefit from personalized treatment strategies and surveillance. The importance of germline

predisposing variants has been proposed with an elevated RIS risk for patients harboring germline *RBI* mutations (50). Genomic analyses from childhood cancer survivors revealed that germline mutations in DNA repair pathway may be exacerbated by radiation therapy and hence contribute to the risk of secondary malignancy (51). In this study, we have identified several candidates of predisposing germline variants in terms of DDR pathway including homology-dependent recombination and base excision repair pathways. Since we have observed a substantial contribution of NHEJ in the generation of indels and SVs in RIS genomes, it is assumed that a deficiency of homology-dependent recombination may lead to alternative mechanisms such as NHEJ in repairing DNA double strand breaks associated with ionizing radiation. However, the further investigation will be required for the functional validation of predisposing germline variants along with the roles of frequent variants such as *TDG*, *TCEB1* and *POLD1*.

The most frequently altered gene in RIS genomes was *MYC* amplification occurring in more than half of the RIS genomes and recurrent focal SCNAs were observed on known cancer-related genes including *NF1*, *RBI* and *CDKN2A*. Recurrent *MYC* alterations were also identified in previous studies, and the frequency was higher than the sporadic sarcomas (52). Mutations and SVs may involve cancer-related genes, but to lesser extent in cohort frequency. In our cohort, *NF1* and *NF2* mutations were truncating ones (i.e., one nonsense and one frameshifting, respectively) likely representing inactivating events of given cases. Although *NF1* and *NF2* mutations were singletons, these genes were further found to harbor additional recurrent SCNAs and SVs. Thus, the genetic alterations in RIS genomes show the evidence of functional convergence in terms of their affecting types of genetic alterations also highlighting that relevance of versatile screening tools such as WGS.

Along with *RAD51B*, the recurrent targets of SVs with roles in homologous

recombination pathway in a haploinsufficient manner (53). Combined with other genomic alterations, this highlights the importance of DDR pathways in RIS genomes. Moreover, complex chromosomal rearrangement, *chromothripsis*, was identified in RIS samples. This also can be a major mutational pathway driving radiation-induced secondary malignancy, and further study is warranted (21).

## **Conclusion**

The current study presents the whole genome-scaled landscape of genetic alterations in RIS, and demonstrates that HR deficiency and NHEJ repair process were important factors for development of RIS. This may serve as a resource to develop diagnostic markers and therapeutic targets for overcoming unresectable RIS.

## References

1. Wakeford, R. (2004) The cancer epidemiology of radiation. *Oncogene*, **23**, 6404-6428.
2. Inskip, P.D., Sigurdson, A.J., Veiga, L., Bhatti, P., Ronckers, C., Rajaraman, P., Boukheris, H., Stovall, M., Smith, S., Hammond, S. *et al.* (2016) Radiation-Related New Primary Solid Cancers in the Childhood Cancer Survivor Study: Comparative Radiation Dose Response and Modification of Treatment Effects. *Int J Radiat Oncol Biol Phys*, **94**, 800-807.
3. Ravanat, J.L., Breton, J., Douki, T., Gasparutto, D., Grand, A., Rachidi, W. and Sauvaigo, S. (2014) Radiation-mediated formation of complex damage to DNA: a chemical aspect overview. *Br J Radiol*, **87**, 20130715.
4. Gonin-Laurent, N., Gibaud, A., Huygue, M., Lefevre, S.H., Le Bras, M., Chauveinc, L., Sastre-Garau, X., Doz, F., Lumbroso, L., Chevillard, S. *et al.* (2006) Specific TP53 mutation pattern in radiation-induced sarcomas. *Carcinogenesis*, **27**, 1266-1272.
5. Hadj-Hamou, N.S., Ugolin, N., Ory, C., Britzen-Laurent, N., Sastre-Garau, X., Chevillard, S. and Malfoy, B. (2011) A transcriptome signature distinguished sporadic from postradiotherapy radiation-induced sarcomas. *Carcinogenesis*, **32**, 929-934.
6. Consortium, I.T.P.-C.A.o.W.G. (2020) Pan-cancer analysis of whole genomes. *Nature*, **578**, 82-93.

7. Behjati, S., Gundem, G., Wedge, D.C., Roberts, N.D., Tarpey, P.S., Cooke, S.L., Van Loo, P., Alexandrov, L.B., Ramakrishna, M., Davies, H. *et al.* (2016) Mutational signatures of ionizing radiation in second malignancies. *Nat Commun*, **7**, 12605.
8. Davidson, P.R., Sherborne, A.L., Taylor, B., Nakamura, A.O. and Nakamura, J.L. (2017) A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies. *Sci Rep*, **7**, 7645.
9. Cahan, W.J.C. (1948) Sarcoma arising in irradiated bone: report of eleven cases. **1**, 3-29.
10. Kim, M., Rhee, J.K., Choi, H., Kwon, A., Kim, J., Lee, G.D., Jekarl, D.W., Lee, S., Kim, Y. and Kim, T.M. (2017) Passage-dependent accumulation of somatic mutations in mesenchymal stromal cells during in vitro culture revealed by whole genome sequencing. *Sci Rep*, **7**, 14508.
11. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
12. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491-498.
13. Jones, D., Raine, K.M., Davies, H., Tarpey, P.S., Butler, A.P., Teague, J.W., Nik-Zainal, S. and Campbell, P.J. (2016) cgpCaVEManWrapper: Simple

- Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr Protoc Bioinformatics*, **56**, 15 10 11-15 10 18.
14. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865-2871.
  15. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, **38**, e164.
  16. Raine, K.M., Van Loo, P., Wedge, D.C., Jones, D., Menzies, A., Butler, A.P., Teague, J.W., Tarpey, P., Nik-Zainal, S. and Campbell, P.J. (2016) ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr Protoc Bioinformatics*, **56**, 15 19 11-15 19 17.
  17. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P.J.N.b. (2011) Integrative genomics viewer. **29**, 24-26.
  18. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R. and Getz, G. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*, **12**, R41.
  19. Gu, Z., Gu, L., Eils, R., Schlesner, M. and Brors, B.J.B. (2014) circlize implements and enhances circular visualization in R. **30**, 2811-2812.

20. Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S. *et al.* (2018) Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep*, **23**, 239-254 e236.
21. Cortes-Ciriano, I., Lee, J.J., Xi, R., Jain, D., Jung, Y.L., Yang, L., Gordenin, D., Klimczak, L.J., Zhang, C.Z., Pellman, D.S. *et al.* (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nature genetics*, **52**, 331-341.
22. Lee, D.D. and Seung, H.S. (1999) Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788-791.
23. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S. and Swanton, C. (2016) DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*, **17**, 31.
24. Wang, S., Li, H., Song, M., He, Z., Wu, T., Wang, X., Tao, Z., Wu, K. and Liu, X.-S.J.m. (2020) Copy number signature analyses in prostate cancer reveal distinct etiologies and clinical outcomes.
25. Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M. *et al.* (2015) Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, **350**, 207-211.
26. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L. *et al.*



- (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415-421.
27. Polak, P., Kim, J., Braunstein, L.Z., Karlic, R., Haradhavala, N.J., Tiao, G., Rosebrock, D., Livitz, D., Kubler, K., Mouw, K.W. *et al.* (2017) A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nature genetics*, **49**, 1476-1486.
28. Rose Li, Y., Halliwill, K.D., Adams, C.J., Iyer, V., Riva, L., Mamunur, R., Jen, K.Y., Del Rosario, R., Fredlund, E., Hirst, G. *et al.* (2020) Mutational signatures in tumours induced by high and low energy radiation in Trp53 deficient mice. *Nat Commun*, **11**, 394.
29. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.
30. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I. and Forbes, S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*, **18**, 696-705.
31. Mani, M., Carrasco, D.E., Zhang, Y., Takada, K., Gatt, M.E., Dutta-Simmons, J., Ikeda, H., Diaz-Griffero, F., Pena-Cruz, V., Bertagnolli, M. *et al.* (2009) BCL9 promotes tumor progression by conferring enhanced proliferative, metastatic, and angiogenic properties to cancer cells. *Cancer Res*, **69**, 7577-7586.
32. Agrawal, N., Frederick, M.J., Pickering, C.R., Bettegowda, C., Chang, K., Li, R.J., Fakhry, C., Xie, T.X., Zhang, J., Wang, J. *et al.* (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating

- mutations in NOTCH1. *Science*, **333**, 1154-1157.
33. Francis, P., Namlos, H.M., Muller, C., Eden, P., Fernebro, J., Berner, J.M., Bjerkehaugen, B., Akerman, M., Bendahl, P.O., Isinger, A. *et al.* (2007) Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. *BMC Genomics*, **8**, 73.
  34. Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B. *et al.* (2016) Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol*, **34**, 155-163.
  35. Merritt, N.M., Fullenkamp, C.A., Hall, S.L., Qian, Q., Desai, C., Thomason, J., Lambertz, A.M., Dupuy, A.J., Darbro, B. and Tanas, M.R. (2018) A comprehensive evaluation of Hippo pathway silencing in sarcomas. *Oncotarget*, **9**, 31620-31636.
  36. Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R. and Campbell, P.J. (2017) Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, **171**, 1029-1041 e1021.
  37. Blazer, L.L., Lima-Fernandes, E., Gibson, E., Eram, M.S., Loppnau, P., Arrowsmith, C.H., Schapira, M. and Vedadi, M. (2016) PR Domain-containing Protein 7 (PRDM7) Is a Histone 3 Lysine 4 Trimethyltransferase. *J Biol Chem*, **291**, 13509-13519.
  38. Hohenauer, T. and Moore, A.W. (2012) The Prdm family: expanding roles

- in stem cells and development. *Development*, **139**, 2267-2282.
39. Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I. *et al.* (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature genetics*, **45**, 136-144.
  40. Briggs, S. and Tomlinson, I.J.T.J.o.p. (2013) Germline and somatic polymerase  $\epsilon$  and  $\delta$  mutations define a new class of hypermutated colorectal and endometrial cancers. **230**, 148-153.
  41. Occidental, M., Shen, G., Feng, X., Zhu, K., Kelly, K., Nie, Q., Reddi, H.V., Lakiotaki, E., Viniou, N.A., Korkolopoulou, P. *et al.* (2020) Novel CTNND2-TERT fusion in a spindle cell liposarcoma. *Genes Chromosomes Cancer*, **59**, 544-548.
  42. Pradhan, B., Sarvilinna, N., Matilainen, J., Aska, E., Sjoberg, J. and Kauppi, L. (2016) Detection and screening of chromosomal rearrangements in uterine leiomyomas by long-distance inverse PCR. *Genes Chromosomes Cancer*, **55**, 215-226.
  43. Golmard, L., Caux-Moncoutier, V., Davy, G., Al Ageeli, E., Poirot, B., Tirapo, C., Michaux, D., Barbaroux, C., d'Enghien, C.D., Nicolas, A. *et al.* (2013) Germline mutation in the RAD51B gene confers predisposition to breast cancer. *BMC Cancer*, **13**, 484.
  44. Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E. and Sfeir, A.J.N. (2015) Mammalian polymerase  $\theta$  promotes alternative

- NHEJ and suppresses recombination. **518**, 254-257.
45. Cai, H., Kumar, N., Bagheri, H.C., von Mering, C., Robinson, M.D. and Baudis, M. (2014) Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *BMC Genomics*, **15**, 82.
  46. Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A. and Begthel, H.J.C. (2018) A living biobank of breast cancer organoids captures disease heterogeneity. **172**, 373-386. e310.
  47. Li, H., Tu, J., Zhao, Z., Chen, L., Qu, Y., Li, H., Yao, H., Wang, X., Lee, D.-F. and Shen, J.J.T. (2020) Molecular signatures of BRCAness analysis identifies PARP inhibitor Niraparib as a novel targeted therapeutic strategy for soft tissue Sarcomas. **10**, 9477.
  48. Gulhan, D.C., Lee, J.J.-K., Melloni, G.E., Cortes-Ciriano, I. and Park, P.J.J.N.g. (2019) Detecting the mutational signature of homologous recombination deficiency in clinical samples. **51**, 912-919.
  49. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A. and Bergstrom, E.N.J.N. (2020) The repertoire of mutational signatures in human cancer. **578**, 94-101.
  50. Klei, R.A., Tucker, M.A., Tarone, R.E., Abramson, D.H., Seddon, J.M., Stovall, M., Li, F.P. and Fraumeni, J.F., Jr. (2005) Risk of new cancers after radiotherapy in long-term survivors of retinoblastoma: an extended follow-up. *J Clin Oncol*, **23**, 2272-2279.

51. Qin, N., Wang, Z., Liu, Q., Song, N., Wilson, C.L., Ehrhardt, M.J., Shelton, K., Easton, J., Mulder, H. and Kennetz, D. (2019) Pathogenic Germline Mutations in DNA-Repair Genes in Combination with Cancer Treatment Exposures and Risk of Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer.
52. Lesluyes, T., Baud, J., Perot, G., Charon-Barra, C., You, A., Valo, I., Bazille, C., Mishellany, F., Leroux, A. and Renard-Oldrini, S.J.M.P. (2019) Genomic and transcriptomic comparison of post-radiation versus sporadic sarcomas. **32**, 1786-1794.
53. Date, O., Katsura, M., Ishida, M., Yoshihara, T., Kinomura, A., Sueda, T. and Miyagawa, K. (2006) Haploinsufficiency of RAD51B causes centrosome fragmentation and aneuploidy in human cells. *Cancer Res*, **66**, 6018-6024.

# 전장 유전체 분석을 이용한 방사선 유발 육종의 유전적 특성 규명

김은지

서울대학교 대학원

임상의과학과

**배경 및 목적:** 방사선 유발 육종은 전리 방사선 치료 후 일정 기간이 지난 후 발생하는 희귀한 2차 악성 종양이다. 안 좋은 임상 예후에도 불구하고, 방사선 유발 육종 발생과 관련한 유전적 기전은 거의 알려지지 않았다. 이 연구는 방사선 유발 육종의 돌연변이 특성을 탐구하기 위해 방사선 유발 육종으로 진단된 환자로부터 얻은 샘플을 통해 전장 유전체 서열을 분석하였다.

**방법:** 2000년과 2019년 사이에 진단된 방사선 유발 육종 환자를 검토하였다. 이전에 방사선 조사를 받은 부위에서 발생한 30개의 육종을 경험이 풍부한 두 명의 병리학자가 검토하였다. 동일한 개체에서 추출한 정상과 종양 조직에서 DNA를 얻었다. 성공적으로 라이브러리

생성이 완료된 11개의 샘플을 최종적으로 포함시켰으며, 종양과 정상조직에 각각 90 x, 60 x 깊이로 염기 서열 분석을 시행하였다. 단일 염기 서열 변이, 짧은 삽입/결실, 복제 수 변화, 구조 변이 및 생식 돌연변이 등을 분석하기 위한 파이프라인을 구축하였다.

**결과:** 방사선 유발 육종 유전체 중 하나는 과돌연변이를 보였으며, 다양한 정도를 보였다. 산화성 염기 변화와 A·T → G·C 전이를 보이는 돌연변이 특성은 방사선 유발 육종에서 흔히 나타났다. 암 관련 유전자 중 *NF1*, *NF2*, *NOTCH1*, *NOTCH2*, *PIK3CA*, *RBI*, *TP53*에서는 단일 체세포 돌연변이가, *MYC*, *CDKN2A*, *RBI*, *NF1*에서는 반복 복제 수 변화가 보였으며, *NF2*, *ARID1B*, *RAD51B*에서는 반복 구조 변이가 관찰되는 등 방사선 유발 육종의 유전체 특징이 관찰되었다. 자발적 골육종과 비교하여, 비상동성말단결합의 영향은 방사선유발육종 유전체의 특징인 짧은 삽입-결실 및 구조적 변이에서 상당히 보여진다. 또한, 빈번한 유전체 산산조각 현상 (chromothripsis) 및 유전자 손상 복구 경로에서 생식 세포 돌연변이들이 관찰되었다.

**결론:** 방사선 유발 육종 유전체에서 전장 유전체 분석을 통한 특성을 종합하여 방사선 유발 육종에 대한 발전된 진단 및 치료 전략을 위한 방법을 모색할 수 있을 것이다.

**주요어:** 전장 유전체 분석, 방사선 유발 육종, 방사선, 이차성 종양, 체세포 돌연변이, 생식세포 돌연변이, 구조적 변이

**학 번:** 2015-30808