Cancer development is influenced by hereditary mutations, somatic mutations due to random errors in DNA replication, or external factors. It remains unclear how distinct cell-intrinsic and -extrinsic factors impact oncogenesis within the same tissue type. We investigated murine soft tissue sarcomas generated by oncogenic alterations (Kras\textsuperscript{G12D} activation and p53 deletion), carcinogens (3-methylcholanthrene [MCA] or ionizing radiation), and in a novel model combining both factors (MCA plus p53 deletion). Whole-exome sequencing demonstrated distinct mutational signatures in individual sarcoma cohorts. MCA-induced sarcomas exhibited high mutational burden and predominantly G-to-T transversions, while radiation-induced sarcomas exhibited low mutational burden and a distinct genetic signature characterized by C-to-T transitions. The indel to substitution ratio and amount of gene copy number variations were high for radiation-induced sarcomas. MCA-induced tumors generated on a p53-deficient background showed the highest genomic instability. MCA-induced sarcomas harbored mutations in putative cancer-driver genes that regulate MAPK signaling (\textit{Kras} and \textit{Nf1}) and the Hippo pathway (\textit{Fat1} and \textit{Fat4}). In contrast, radiation-induced sarcomas and Kras\textsuperscript{G12D}p53\textsuperscript{−/−} sarcomas did not harbor recurrent oncogenic mutations, rather they exhibited amplifications of specific oncogenes: \textit{Kras} and \textit{Myc} in Kras\textsuperscript{G12D}p53\textsuperscript{−/−} sarcomas, and \textit{Met} and \textit{Yap1} for radiation-induced sarcomas. These results reveal that different initiating events drive oncogenesis through distinct mechanisms.

Find the latest version:

http://jci.me/128698/pdf
Mutational landscape in genetically engineered, carcinogen-induced, and radiation-induced mouse sarcoma

Chang-Lung Lee¹,²,³*, Yvonne M. Mowery¹,²*, Andrea R. Daniel¹*, Dadong Zhang², Alexander B. Sibley², Joe R. Delaney⁴, Amy J. Wisdom⁵, Xiaodi Qin², Xi Wang², Isibel Caraballo¹, Jeremy Gresham², Lixia Luo¹, David Van Mater²,⁶, Kouros Owzar²,⁷ and David G. Kirsch¹,²,⁵

¹Department of Radiation Oncology, ²Duke Cancer Institute, ³Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA. ⁴Department of Biochemistry and Molecular Biology, Medical University of South Carolina (MUSC), Charleston, SC 29425, USA. ⁵Department of Pharmacology & Cancer Biology, ⁶Department of Pediatrics, and ⁷Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA.

*Equal contribution

CORRESPONDENCE

David G. Kirsch, Duke University Medical Center, Box 91006, Durham, NC 27708, Phone: 919-681-8605, Fax: 919-681-1867, E-mail: david.kirsch@duke.edu

CONFLICT OF INTEREST

The authors have no conflicting financial interests.
ABSTRACT

Cancer development is influenced by hereditary mutations, somatic mutations due to random errors in DNA replication, or external factors. It remains unclear how distinct cell-intrinsic and -extrinsic factors impact oncogenesis within the same tissue type. We investigated murine soft tissue sarcomas generated by oncogenic alterations (Kras^{G12D} activation and p53 deletion), carcinogens (3-methylcholanthrene [MCA] or ionizing radiation), and in a novel model combining both factors (MCA plus p53 deletion). Whole-exome sequencing demonstrated distinct mutational signatures in individual sarcoma cohorts. MCA-induced sarcomas exhibited high mutational burden and predominantly G-to-T transversions, while radiation-induced sarcomas exhibited low mutational burden and a distinct genetic signature characterized by C-to-T transitions. The indel to substitution ratio and amount of gene copy number variations were high for radiation-induced sarcomas. MCA-induced tumors generated on a p53-deficient background showed the highest genomic instability. MCA-induced sarcomas harbored mutations in putative cancer-driver genes that regulate MAPK signaling (Kras and Nf1) and the Hippo pathway (Fat1 and Fat4). In contrast, radiation-induced sarcomas and Kras^{G12D}p53^{-/-} sarcomas did not harbor recurrent oncogenic mutations, rather they exhibited amplifications of specific oncogenes: Kras and Myc in Kras^{G12D}p53^{-/-} sarcomas, and Met and Yap1 for radiation-induced sarcomas. These results reveal that different initiating events drive oncogenesis through distinct mechanisms.
INTRODUCTION

Approximately 50% of all cancer patients receive radiation therapy, which is a component of approximately 40% of all cancer cures. While radiation is an effective cancer therapy, its use involves a small, but clinically significant risk of developing a therapy-related malignancy. Radiation-associated cancers develop years later and are a particular concern for pediatric cancer patients as they may carry germline mutations in oncogenes or tumor suppressors genes and as they have many years to develop secondary cancers. Moreover, the estimated total lifetime risk of radiation-associated cancers may be higher in patients receiving modern radiation therapy techniques, such as intensity modulated radiation therapy (IMRT) and image guided radiation therapy (IGRT). When a second cancer develops after radiation exposure, it can be challenging to determine whether radiation caused the tumor.

Radiotherapy kills cells by generating unresolved double stranded DNA breaks. For example, cells undergo mitosis with unrepaired double stranded DNA breaks after radiotherapy can die through mechanisms including mitotic catastrophe. While radiation effectively kills proliferating cancer cells, it is a relatively weak carcinogen. In contrast, the potent chemical carcinogen 3-methylcholanthrene (MCA) is non-lethal, but acts as a mutagen to modify DNA sequences, primarily causing G to T transversions. Mutagenesis initiates a selection process that favors proliferative cells harboring activated oncogenes and inactivated tumor suppressor genes. However, mechanisms by which radiation-induced DNA damage and repair processes cause de novo cancer formation, as well as the specific types of DNA mutations and pathways modulated, remain poorly understood.

To identify mutational signatures specific to radiation-induced tumors and to gain insight into how distinct cell-intrinsic and -extrinsic factors impact cancer development within the same
tissue type, we performed genomic analysis across murine soft-tissue sarcomas induced by MCA, oncogenic mutations, or ionizing radiation. Radiation-induced sarcomas were generated by focally irradiating the mouse hind limb using a single dose of 30-40 Gy(9). For comparison to radiation-induced sarcomas, we utilized an established genetically engineered mouse model of soft tissue sarcoma in which localized delivery of Cre recombinase into the muscle of the hind limb activates oncogenic \(Kras^{G12D}\) and deletes both alleles of \(p53\)(10). In addition, we generated MCA-induced sarcomas in the hind limb of either WT or \(p53^{FL/FL}\) mice in which both copies of \(p53\) were deleted by Cre recombinase.

Using these mouse models of oncogene-driven, chemical carcinogen-induced or radiation-induced soft tissue sarcoma, we performed whole-exome sequencing (WES) on paired tumor and normal tissue from each mouse and observed distinct facultative molecular signatures that are specific to each carcinogenic driver. Remarkably, ionizing radiation produced tumors with relatively low levels of nonsynonymous mutations, but a high frequency of somatic copy-number alterations, with a preponderance of deletions and a predilection for C to T and G to A transitions.

**RESULTS**

**Generation of primary murine sarcomas by oncogenic alterations, chemical carcinogens, and ionizing radiation**

To investigate how cell-intrinsic and -extrinsic factors impact cancer development within the same tissue type, we generated primary murine sarcomas by using defined genetic and external insults including mutations of Kras and \(p53\)(10) (\(Kras^{G12D}\) \(p53^{-/-}\)), chemical carcinogen MCA(11) (MCA-induced \(p53\) WT and MCA-induced \(p53^{-/-}\)) and ionizing radiation(9) (IR-induced)
(Figure 1a and Table S1). Kras\textsuperscript{G12D} p53\textsuperscript{−/−} sarcomas were generated in LSL-Kras\textsuperscript{G12D} \(; p53^{FL/FL}\) mice following intramuscular delivery of adenovirus expressing Cre recombinase (Ad-Cre)(10). The median of the observed event times for the six mice with Kras\textsuperscript{G12D} p53\textsuperscript{−/−} sarcomas for which data were available was 77 days after Ad-Cre injection. The latency of Kras\textsuperscript{G12D} p53\textsuperscript{−/−} sarcomas was similar to MCA-induced p53\textsuperscript{−/−} sarcomas, which were generated via intramuscular injection of both Ad-Cre and MCA into \(p53^{FL/FL}\) mice (Table S2). Compared to MCA-induced p53\textsuperscript{−/−} and Kras\textsuperscript{G12D} p53\textsuperscript{−/−} sarcomas, MCA-induced sarcomas generated in p53 WT mice had markedly longer latency (Table S2). Notably, IR-induced sarcomas, which developed in mice with or without temporary (10 day) p53 knockdown during a single dose of 30 or 40 Gy focal irradiation(12), had the longest observed latency. The median observed event time was 449 days after radiation exposure (Table S2). Histology demonstrated that sarcomas generated by these approaches were intermediate- to high-grade soft tissue sarcomas (Figure S1). Collectively, these mouse models provide a unique resource to comprehensively understand the mutational landscape across soft tissue sarcomas generated through distinct oncogenic alterations and carcinogens.

**Tumor-initiating factors dictate mutational load and signatures**

To determine the mutational landscape of soft tissue sarcomas, we performed WES on paired tumor and normal liver to identify somatic mutations that are specific to each tumor model (Figure S2). Comparing across different sarcoma cohorts, MCA-induced sarcomas harbored the highest mutational burden (Figure 1, b and c). Both MCA-induced p53 WT and MCA-induced p53\textsuperscript{−/−} sarcomas contain a median of > 2,000 nonsynonymous mutations per tumor (Figure 1c). IR-induced sarcomas harbored a substantially lower mutational load with a median of 26
nonsynonymous mutations per tumor. The mutational burden of IR-induced sarcomas was similar to Kras$^{G12D}$ p53$^{-/-}$ sarcomas (Figure 1, b and c). Notably, a single IR-induced sarcoma (S28) exhibited a disproportionately high number of mutations, of which about 15% were localized on chromosome 2 (Figure S3). Further examination of S28 revealed mutations in multiple genes that control the DNA damage response, including Brca1, Atrx, and Pole (Table S3), suggesting that defects in DNA repair and cell cycle checkpoint controls lead to an accumulation of mutations in this tumor(13). Together, these results indicate that while MCA generates sarcomas by causing gene mutations, ionizing radiation does not typically induce sarcomas by increasing mutational burden.

In addition to assessing the number of mutations, we examined the impact of different genetic and external insults on the distribution of sequence variants, including single-nucleotide variants (SNVs) and insertions-deletions (indels) (Figure 1d and Figure S4). Compared to MCA-induced sarcomas and Kras$^{G12D}$ p53$^{-/-}$ sarcomas, IR-induced sarcomas showed a higher median proportion of nonsynonymous mutations that are indels (p=0.0003, Figure 1d). Further investigation revealed that nonsynonymous indels in IR-induced sarcomas were predominately deletions (Figure 1e). Moreover, examination of SNVs showed that while MCA-induced sarcomas harbored primarily C to A and G to T transversions, IR-induced sarcomas exhibited higher C to T (p=0.0002) and G to A (p=0.0006) transitions (Figure 1f). The distinction of single-nucleotide substitutions was also revealed by unsupervised hierarchical clustering showing segregation between the majority of MCA-induced sarcomas, Kras$^{G12D}$ p53$^{-/-}$ sarcomas and IR-induced sarcomas (Figure 1g).

We further conducted signature analysis using nonnegative matrix factorization (NMF)(14), and compared our results to 30 published signatures identified in human cancers
Our results revealed that mutational signatures derived from each murine sarcoma cohort were highly correlated with distinct COSMIC human signatures (Figure S5). COSMIC signature 4, which is associated with tobacco mutagens, was exclusively enriched in MCA-induced sarcomas (Figure S5, a and b). COSMIC signature 5, which is present universally in all 30 types of human cancers(15), was enriched in a subset of Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas and IR-induced sarcomas. While COSMIC signatures 9 and 17 were specifically found in certain Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas, IR-induced sarcomas exhibited a signature that correlated with COSMIC signature 6, which may indicate microsatellite instability (Figure S5, c and d). In sum, as shown in the mutational analysis, our results reveal unique mutational processes underlying the development of sarcomas induced by Kras and p53 mutations, MCA carcinogen, and ionizing radiation.

**Ionizing radiation and p53 status contribute to increased copy number variations**

In addition to examining mutations, we evaluated somatic copy number variations (CNVs) using CODEX2(16) (Figure S2, Figure 2a and Figure S6). Among sarcomas initiated in p53 WT mice, IR-induced sarcomas exhibited a markedly higher median number of genes affected by CNVs compared to MCA-induced p53 WT sarcomas (p=0.0262, Figure 2b). This trend was consistent for both copy number gains and losses (p=0.0262 and 0.0297, Figure 2, c and d, respectively). Moreover, sarcomas induced by MCA in p53 WT mice showed a lower median number of genes affected by CNVs compared to sarcomas initiated by MCA and p53 loss, suggesting that the p53 status of tumor cells either at the time of MCA exposure or during subsequent tumor development had a marked impact on chromosomal instability (Figure 2, b to d). Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas, which did not develop after an external genotoxic exposure,
showed a similar median number of genes with CNVs as IR-induced sarcomas (Figure 2, b to d). Together, these findings suggest that both ionizing radiation and p53 loss contribute to increase the number of CNVs during sarcomagenesis.

**Different sarcoma cohorts show enrichment in genes affected by mutations versus CNVs**

To elucidate genetic alterations that contribute to sarcoma development, we compared the number of genes affected by mutations versus the number affected by CNVs in each sarcoma sample. Both IR-induced sarcomas and Kras^{G12D} p53^{-/-} sarcomas were defined by a markedly higher number of genes affected by CNVs than mutations (Figure 3, a and b). In contrast, the majority of MCA-induced tumors exhibited relatively few genes affected by CNVs compared to mutations (Figure 3, a and b). Of note, about 50% of MCA-induced p53^{-/-} sarcomas were clustered at the top right of the graph due to harboring both nonsynonymous SNVs and CNVs in high number of genes (Figure 3a).

To examine genetic alterations that potentially contribute to oncogenesis, we evaluated specific oncogenic genes that were impacted by mutations and CNVs using the COSMIC database. While the number of nonsynonymous mutations in COSMIC genes was extremely low in radiation-induced and Kras^{G12D} p53^{-/-} sarcomas, frequent mutations were observed in COSMIC genes in the MCA-induced tumors (Figure 3c). In contrast, the median number of COSMIC genes affected by CNVs was higher in IR-induced sarcomas, Kras^{G12D} p53^{-/-} sarcomas and MCA-induced p53^{-/-} sarcomas compared to MCA-induced p53 WT sarcomas (Figure 3d).

**Mutations in putative driver genes of sarcomas**
To evaluate putative driver genes, we analyzed recurring nonsynonymous mutations and CNVs of COSMIC genes in different sarcoma cohorts. Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas showed essentially no recurring mutations in COSMIC genes. IR-induced sarcomas harbored recurring mutations in only four COSMIC genes, despite the analysis including the hypermutated sample S28 (Figure 4a). In contrast, MCA-induced sarcomas exhibited a high frequency of mutations in numerous putative driver genes, including Kras and NF1 (Figure 4b). Of note, we observed p53 mutations in 100% of sarcomas (6 out of 6) that developed from p53 WT mice treated with MCA. The majority of these p53 mutations were missense mutations located in the DNA binding domain (Figure S7). However, no p53 mutations were observed in sarcomas (0 out of 8) that developed in p53 WT mice induced by ionizing radiation.

Examination of genes affected by CNVs revealed amplification of a distinct spectrum of COSMIC oncogenes in Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas versus IR-induced sarcomas (Figure 5a, Table S4 to S7). While a subset of Kras\textsuperscript{G12D} p53\textsuperscript{-/-} sarcomas had amplifications of oncogenes Kras and Myc, several IR-induced sarcomas exhibited prominent amplifications of Met and Birc3 (Figure 5a). An increase in copy number of Met and Birc3 resulted from partial amplifications of Chromosomes 6 and 9, respectively (Figure 5b). The fragment that was amplified on Chromosome 9 contains multiple putative driver genes, including Yap1 (Table S4). To validate the results from the WES data, we performed qRT-PCR to examine CNVs of Met, Birc3 and Yap1 (Figure 5c). Our results from qRT-PCR were consistent with the findings from WES, showing amplification of Met in IR-induced sarcomas S28, S31, S32, as well as amplifications of Birc3 and Yap1 in IR-induced sarcomas S32 and S33.

DISCUSSION
Radiation-associated sarcomas are a rare but significant potential late side effect of radiation therapy(17). However, methods are currently lacking to discern whether a second malignancy is caused by radiation exposure. To date, a robust mutational signature for distinguishing ionizing radiation-initiated cancers from tumors driven by other pathogenetic events has not been defined. Because the genetic drivers of radiation-related cancers may differ from spontaneous cancers, identifying specific genetic features in tumors that contribute to an ionizing radiation signature has the potential to not only to impact diagnosis, but also therapy. Searching for a genetic signature of radiation-associated cancer in human samples is complicated by variations in radiation dose and fractionation, anatomic location, tumor type, and uncertainty regarding whether radiation initiated the tumor. By contrast, our primary murine sarcoma models provide a well-controlled system to search for a genetic signature of radiation-driven tumorigenesis. We employed WES to characterize the genetic changes in sarcomas derived from four mouse models with distinct and clearly defined tumor initiating events: high dose focal ionizing radiation, chemical carcinogen (MCA), p53 loss with a chemical carcinogen (MCA), and p53 loss with Kras activation.

A mutational signature depends on the mechanism of mutagenesis and subsequent selection process that malignant cells undergo during tumor development. For example, MCA metabolites form covalent bonds with double- and single-stranded DNA, preferentially at guanine residues to produce G to T transversions(8). Therefore, the specific base changes that predominate in the MCA-induced p53 WT and MCA-induced p53−/− sarcomas are G to T and the reverse (C to A) single base substitutions (Figure 1f). Ionizing radiation generates DNA damage when energy is directly absorbed by DNA molecules and indirectly through ionization of water or other intracellular molecules to generate hydroxyl radicals that cause 2-deoxyribose
oxidation(18, 19). Guanine residues are particularly sensitive to oxidation compared to cytosine, thymine, and adenine, and 8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxoG) is among the most readily detected base products after ionizing radiation(20). Subsequently, 8-oxoG itself is far more susceptible to further oxidation, yielding more stable molecules including spiroiminodihydantion and guanidinohydantoin which are more mutagenic(18). 8-oxoG adducts predominately lead to G:C to T:A transitions(21). Furthermore, reactive oxygen species through Fenton chemistry lead to deamination of methylated cytosines and thymine single base substitutions(22, 23). Therefore, G to A and C to T DNA transition mutations are hallmarks of oxidative damage(23). Previous analysis of radiation-associated human tumors(24, 25) and radiation-induced mouse tumors(26) reported a prevalence of C to T transitions. Our data, which include specific controls for alternative tumor initiating events, demonstrate a preference for C to T and the reverse (G to A) base sequence mutations in radiation-induced tumors (Figure 1f), indicating a strong oxidative mutation signature generated by ionizing radiation.

While the single base substitution patterns for each tumor model reveal distinguishing underlying mechanistic information, the overall somatic mutational load also provides insights into tumor initiation. MCA is a potent mutagen that generates sarcomas with roughly 80 times the median number of mutations compared to IR-induced sarcomas or KrasG12D p53/+ sarcomas (Figure 1b). Thus, the MCA-driven p53/+ sarcomas represent a novel spatially- and temporally-restricted high mutational load mouse model in which autochthonous tumors develop over 10-18 weeks (Table S2), evolving under the selective pressure of an intact immune system. In contrast to conventional genetically engineered mouse models, such as the KrasG12D p53/+ sarcomas, the MCA-driven p53/+ sarcomas exhibit a mutational load similar to many human cancers that respond to immunotherapy(27, 28). Therefore, this model will be an important new tool to study
the co-evolution of tumors with the immune system and a pre-clinical platform to test immunotherapy.

Remarkably, the radiation-induced sarcomas exhibited relatively few nonsynonymous somatic mutations (Figure 1c). The low mutational load in the radiation-induced tumors is surprising, but this is consistent with radiation acting as a relatively weak carcinogen(3, 7). Notably, others have reported higher mutational loads in radiation-induced mouse tumors(26). Potential explanations for this discrepancy include differences in tumor types analyzed, radiation dose and fractionation. Moreover, a reference genome was used to call somatic mutations(26), which has the potential to increase the number of called mutations. In contrast, we performed WES using paired normal tissue as the reference for each tumor. Consistent with our findings, studies examining human radiation-associated tumors reported a relatively low mutational load(24, 25, 29). Because we did not sequence other tumor types or include tumors that developed after fractionated radiation exposure, the signature defined herein may not be universal for all radiation-induced cancers.

Although radiation-induced sarcomas exhibited a low number of mutations, they were comprised of a higher proportion of nonsynonymous deletion events compared to Kras^{G12D} p53^{-/-} and MCA-driven sarcomas (Figure 1e). This result corroborates the finding from Behjati et al.(25), showing that insertions and deletions were not equally represented in human radiation-associated second malignancies, but rather that deletions were enriched and evenly distributed throughout the genome(25). Although deletions were relatively common, perhaps suggesting a loss of tumor suppressor function, none of the radiation-induced sarcomas in this study exhibited a mutation in the p53 gene. In fact, no specific driver mutations were identified in this tumor cohort (Figure 4a). However, the low mutational burden observed by WES in the mouse
radiation-induced sarcomas represents a limitation for identifying specific driver mutations and conducting NMF signature analysis. In contrast, gene copy number changes were more abundant in radiation-induced sarcomas compared to Kras$^{G12D}$ p53$^{−/−}$ and MCA-driven tumors. Oncogenes Met, Yap1 and Birc3 each exhibited copy number gains in approximately half of the radiation-induced tumors (Figure 5c). Notably, the Yap1 pathway is commonly activated in rhabdomyosarcomas and Yap1 overexpression in muscle satellite cells is sufficient to induce sarcomagenesis in the context of muscle injury(30).

In contrast to the radiation-induced tumors which retained wild type p53 genes, all MCA-induced tumors from wild type mice acquired a p53 mutation (Figure 4b and Figure S7). Although 7 of 8 of the radiation-induced tumors arose from mice that received 10 days of doxycycline to induce p53 shRNA during radiation (i.e. temporary p53 knockdown), doxycycline was removed immediately following irradiation and mice subsequently remained on normal chow for the remainder of the experiment. Notably, the radiation-induced tumor that arose from a mouse lacking the p53 shRNA gene likewise did not harbor a detectable p53 mutation. The MCA-induced tumors that developed on a p53 WT background exhibited increased incidence of tumor suppressor mutations compared to MCA-induced tumors that developed in the setting of Cre-mediated p53 deletion. Interestingly, the MCA-induced p53 WT tumor mutational spectrum differed substantially from that of the MCA-induced p53$^{−/−}$ tumors. The tumors that arose in wild type mice with initially intact p53 developed over a longer period of time and activated different pathways. Indeed, oncogenes Abl2 and Bcl9 and tumor suppressors Nbn, Ptprc, Brca1, and Neor1 were altered in over half of the MCA-induced p53 WT tumors versus almost none of the MCA-induced p53$^{−/−}$ tumors (Figure 4b). These findings suggest that p53 mutation timing, prior to versus as a consequence of MCA exposure, shapes the
mutational landscape by altering the selective pressure for cells to mutate specific genes. Notably, Kras was mutated in half of all MCA-driven tumors independent of p53 status. Moreover, the tumor suppressor Fat1 was mutated in nearly all MCA tumors, and Fat4, Notch2 and NF1 were also commonly disrupted (Figure 4b). Our study comports with sequencing data from a commonly used MCA-driven sarcoma cell line derived from immunodeficient mice (Rag2\^+\^-)(31). Furthermore, our comprehensive analysis of a large cohort of MCA-driven tumors supports the utility of this well characterized primary mouse model of sarcoma for preclinical drug development studies in the presence of an intact immune system.

The genetic landscape of radiation-induced tumors reported here is distinct form published signatures for other carcinogenic processes, such as aging(32) or UV exposure(33). In studies examining radiation-associated liver tumors, higher radiation dose resulted in an increased fraction of cells harboring p53 mutations, likely through a clonal expansion mechanism(34). We previously published a report detailing the non-cell-autonomous mechanism by which radiation induces lymphomagenesis(12). In this case, total-body irradiation eliminates cells in the bone marrow niche, allowing thymic cells with preexisting oncogenic mutations to expand into a tumor unencumbered by cell competition from the bone marrow. However, the mechanisms for radiation-induced sarcomagenesis may be distinct from radiation-induced lymphomagenesis. The WES provides evidence of radiation-induced oxidative DNA damage and amplification of genes such as Met and Yap1, which are both associated with injury-induced sarcomas(35), suggesting a cell-autonomous mechanism. We suspect that after tumor-initiating cells undergo radiation-induced DNA damage, they begin clonal expansion and develop into a tumor through a selection process shaped by acute and chronically injured surrounding tissue following radiation exposure. The microenvironment of irradiated tissue is characterized by high
levels of inflammatory cells and increased growth factor secretion to stimulate wound healing. Tumors that arise under these conditions are adapted to take advantage of the abundant cytokines in this milieu(36). Therefore, radiation-induced cancer may respond to different therapeutic approaches, including immunotherapy, compared to tumors from the same tissue that develop independent of radiation exposure. Defining a signature of radiation-induced cancer that can identify and characterize these tumors is a critical step toward optimizing treatment for this challenging clinical problem.

MATERIALS AND METHODS

Mouse strains and sarcoma induction
To study IR-induced sarcomas, we used previously described mouse models expressing a doxycycline-inducible shRNA against p53 including CMV-rtTA; TRE-p53.1224 and Actin-rtTA; TRE-p53.1224 mice as well as their littermates that only express rtTA or TRE-p53.1224(12). All mice were on a C3H and C57BL/6J mixed genetic background. Six to 24-week-old mice were placed on a doxycycline diet for ten days prior to irradiation(12). The left hind limb of the mice was irradiated with 30 or 40 Gy, and then animals were immediately returned to normal chow. Hind limb irradiation was performed using the X-RAD 225Cx small animal image-guided irradiator (Precision X-Ray). The irradiation field included the whole left hind limb and was defined using fluoroscopy with 40 kVp, 2.5 mA X-rays using a 2 mm Al filter. Irradiations were performed using parallel-opposed anterior and posterior fields with an average dose rate of 300 cGy/min prescribed to midplane with 225 kVp, 13 mA X-rays using a 0.3 mm Cu filter.

Genetically engineered and carcinogen-induced primary sarcomas were generated in 6 to 10-week-old mice with a mixed genetic background. Primary Kras\textsuperscript{G12D}, p53\textsuperscript{-/-} sarcomas were
induced by injection of adenovirus expressing Cre recombinase (Adeno-Cre, Viral Vector Core, University of Iowa) into the gastrocnemius of LSL-Kras\textsuperscript{G12D} \& p53\textsuperscript{FL/FL} mice (10). Carcinogen-induced sarcomas in mice with intact p53 (MCA-induced p53 WT) were generated by intramuscular injection of 300 µg MCA (Sigma-Aldrich) resuspended in sesame oil (Sigma-Aldrich) at 6 µg/µl. MCA-induced sarcomas were induced in the setting of p53 deletion by intramuscular Adeno-Cre injection into the gastrocnemius of p53\textsuperscript{-/-} mice (MCA-induced p53\textsuperscript{-/-}), 24 hours later followed by 300 µg injection of MCA.

After treatment, mice were examined weekly for sarcomas. Upon detection, tumors were harvested with half submerged in RNAlater (ThermoFisher Scientific) for subsequent DNA isolation and half formalin-fixed for histological analysis. Livers were collected for normal tissue control samples.

**WES methods**

Tumor specimens and matched liver control stored in RNAlater were used for DNA extraction. DNA extraction was performed using DNeasy Blood and Tissue Kit or AllPrep DNA/RNA Mini Kit (Qiagen). WES was performed in two batches either using previously described methods (Batch 1)(35) or with the following method (Batch2) (Table S1). One mouse in the Kras\textsuperscript{G12D} p53\textsuperscript{-/-} cohort, S45, was excluded from analyses after whole-exome sequencing showed no evidence of a deletion of p53 exons two to ten. Genomic DNA samples were quantified using fluorometric quantitation on the Qubit 2.0 (ThermoFisher Scientific). For each sample, 200ng of DNA was sheared using focused-ultrasonicators (Covaris) to generate DNA fragments of about 300bp in length. Sequencing libraries were then prepared using the Agilent SureSelect XT Mouse All Exon kit (#S0276129). During adapter ligation, unique indexes were added to each
sample. Resulting libraries were cleaned using Solid Phase Reversible Immobilization (SPRI) beads and quantified on the Qubit 2.0, and size distribution was checked on an Agilent Bioanalyzer. Libraries were subsequently enriched individually by hybridization of the prepared gDNA libraries with mouse all exome target-specific probes provided with the SureSelect XT Mouse All Exon kit. The kit has a target size of 49.6Mb. After hybridization, the targeted molecules are captured on streptavidin beads. Once enriched, the libraries are pooled and sequenced on the Illumina HiSeq 2500 and Illumina HiSeq 4000 with read length of 125bp and 150bp Paired-End sequencing protocol, respectively (Table S8). This pooling scheme generated about 14.5-63.5 million reads per sample or about 6Gb of data. Once generated, sequence data were demultiplexed and Fastq files were generated using Bcl2Fastq2 conversion software provided by Illumina. The sequencing data along with the called mutations in vcf format have been deposited onto NCBI Sequence Read Archive (SRA) under project ID: PRJNA516973.

**WES data analyses**

*Somatic mutation calling:* The raw sequences were first aligned to the mouse reference genome using the BWA-MEM algorithm (v0.7.12-r1039)(37). The reference genome was obtained from Sanger Institute FTP site (ftp://ftp-mouse.sanger.ac.uk/). Reference germline information was obtained from the GATK bundle, somatic mutation information was from COSMIC(38), and SNPs were annotated using SNPeff (39) and Oncotator(40). The original capture file, which had been built on GRCm37 (mm9), was lifted to GRCm38 (mm10) to match with the other reference files. The aligned bam files were preprocessed by using picard-tools (v2.8.3; http://broadinstitute.github.io/picard/faq.html), followed by somatic mutation detection using
GATK3-MuTect2(41). The impact of called mutations was evaluated using Ensembl’s Variant Effect Predictor (VEP) (v91.3) (42) and visualized using R package pheatmap(43).

Somatic mutation plots: Called mutations (SNVs and indels) in GATK3-MuTect2 and annotated by VEP as having “High” or “Moderate” impact were considered “protein-altering”. To determine oncogenic drivers, the COSMIC database(44) was used as a consistent, community-accepted database of tumor suppressors and oncogenes (Release v85). Tier 1 genes were downloaded from the Cancer Gene Census, fusion-only genes removed, and input into MouseMine to determine murine homologues of these oncogenes and tumor suppressor genes. The list of protein-altering mutations was filtered to only those mutations occurring within one of the identified genes, using the Bioconductor(45) R package biomaRt(46) to determine gene locations. If a sample had more than one mutation within a single gene, the mutation of greatest impact was retained. For non-MCA sarcoma samples, genes mutated in two or more samples were included in the figures. For sarcomas induced by MCA, genes mutated in more than 50% of samples in a single tumor type were included.

Mutational signatures: Signature analysis and visualization were conducted using Alexandrov et al.’s method(14) implemented in R package maftools (v1.6.15)(47).

Copy number variation: CNV was analyzed using CODEX2(16) and visualized using R package pheatmap (v1.0.12)(43). Segments of estimated variation were compared to gene positions using the Bioconductor(45) annotation packages TxDb.Mmusculus.UCSC.mm10.knownGene(48) and org.Mm.eg.db(49). If a gene was intersected by more than one segment, the estimated variation
with the longest sequence overlap was retained. Genes with absolute estimated variations greater than 0.2 were considered CNVs. This threshold was determined based on the observed estimated variation of the \( p53 \) gene in samples from \( p53 \) deleted sarcoma cohorts (Figure S8). Genes with CNVs in three or more samples from one tumor cohort were included in the figures.

**qRT-PCR**

Relative genomic DNA levels were determined using qPCR assays performed on the QuantStudio 6 Flex Real-Time PCR System with PowerUp Sybr Green Master Mix (ThermoFisher Scientific, A25742) and specific primer sets designed within exons. Target gene quantification levels were normalized to a housekeeper gene and normal tissue DNA samples using the delta-delta CT method.

*Met* DNA – Forward: AAT ATC CTC CAA GCC GCG TA  
*Met* DNA – Reverse: TGA TGG GGA ATG CAC AGA CT  
*Yap1* DNA – Forward: CAA ATG TGG ACC TTG GCA CA  
*Yap1* DNA – Reverse: CCC TCA CAG ACT CAG AGT GG  
*Brca1* DNA – Forward: CGG ATG CCA AGA AGA ACG AG  
*Brca1* DNA – Reverse: GTT CCT GTT CTC TGA GGG CT  
*Birc3* DNA – Forward: GGA CAG TCC CAT GGA GAA GC  
*Birc3* DNA – Reverse: CAA AGG CAT GGT GCT CAT CG  
*36B4* DNA – Forward: ACT GGT CTA GGA CCC GAG AAG  
*36B4* DNA – Reverse: TCA ATG GTG CCT CTG GAG ATT

**Immunohistochemistry**
Tumor tissue was fixed in 10% neutral buffered formalin for 24-48 hours, preserved in 70% ethanol, and embedded in paraffin. Tissues were sectioned onto a slide and stained with hematoxylin and eosin.

**Statistical considerations**

The p-values presented were two-sided, were the results of post-hoc analyses, and were not adjusted for multiple testing. When comparing a quantitative phenotype with respect to two groups, the Mann-Whitney U test was used, while for three or more groups, the Kruskal-Wallis test was used. All inferential analyses were carried out using the R statistical environment (50) along with extension packages from the comprehensive R archive network (CRAN; https://cran.r-project.org/) and the Bioconductor project (45). Box and whisker plots presented in the figures were constructed as follows: the center line indicated the median value, the bounds represented the first and third quartiles, and the whiskers extended to either a length of 1.5 times the inter-quartile range past the bounds, or the most extreme data value (i.e., minimum or maximum), whichever was shorter. In box and scatter plots, each dot represented the data for one tumor, unless otherwise indicated. In bar plots, each bar represented the data for one tumor, unless otherwise indicated.

**Computational considerations**

The analyses were conducted with adherence to the principles of reproducible analysis using the knitr package (51) for generation of dynamic reports and mercurial (https://www.mercurial-scm.org/) for source code management. The code for replicating the statistical analysis was made

Study approval

All animal procedures for this study were approved by the Institutional Animal Care and Use Committee (IACUC) at Duke University, Durham, NC, USA.

AUTHOR CONTRIBUTIONS

CLL, YMM, ARD, and DGK designed the study. CLL, YMM, ARD, AJW, IC, LL, and DVM performed experiments. CLL, YMM, ARD, DZ, ABS, JRD, AJW, XQ, XW, JG, DVM, KO, DGK analyzed and interpreted data. CLL, YMM, ARD, and DGK wrote the manuscript. All authors edited and approved the manuscript.

ACKNOWLEDGEMENTS

We thank Lorraine da Silva Campos, Nerissa Williams, Yan Ma, and Kennedy Davis Brock for assisting with experiments. This work was supported by the following grants: Department of Defense (W81XWH-18-1-0162 to DGK), National Cancer Institute (R35CA197616 to DGK, K99CA212198 to C-LL and K99CA207729 to JD), Sarcoma Alliance for Research through Collaboration (5U54CA168512-02 to DGK and YM), Radiological Society of North America Research Resident/Fellow Grant (YM), Whitehead Scholar Award from Duke University School of Medicine (C-LL), and the Duke Cancer Center Support Grant 5P30CA14236-44.

REFERENCES


Figure 1. Somatic mutation analysis of murine soft-tissue sarcomas. A, Schematics of the methods to generate various mouse models of soft-tissue sarcomas: IR-induced (blue), MCA-induced p53 WT (red), MCA-induced p53⁻/⁻ (green), and Kras<sup>G12D</sup>, p53⁻/⁻ sarcomas (purple). B, The number of total somatic mutations per tumor. C, The number of somatic nonsynonymous mutations per tumor. D, The proportion of insertion-deletions (indels) within nonsynonymous mutations. IR-induced sarcomas showed a higher median proportion of nonsynonymous mutations that are indels (p=0.0003). E, The proportion of insertions or deletions within nonsynonymous mutations. F, The proportions of different single nucleotide substitutions. IR-induced sarcomas exhibited higher C to T (p=0.0002) and G to A (p=0.0006) transitions. G, Unsupervised hierarchical clustering of sarcomas based on data of single nucleotide substitutions. P-values were calculated by the Kruskal-Wallis test. Panels B-G illustrate the data for n=37 tumors.
Figure 2. Somatic copy number variations CNVs in mouse soft-tissue sarcomas. A, Schematics of CNVs across 19 chromosomes. Results represent pooled data from sarcomas of the same cohort. DNA deletions (del) and duplications (dup) are labeled by blue and red, respectively. B, The number of genes affected by CNVs. IR-induced sarcomas exhibited higher numbers of genes affected by CNVs than MCA-induced p53 WT sarcomas (p=0.0262). C, The number of genes with copy number gains. IR-induced sarcomas exhibited higher numbers of genes with copy number gains than MCA-induced p53 WT sarcomas (p=0.0262). D, The number of genes with copy number losses. IR-induced sarcomas exhibited higher numbers of genes with copy number losses than MCA-induced p53 WT sarcomas (p=0.297). P-values were calculated by the Mann-Whitney U test. Panels illustrate the data for n=37 tumors.
Figure 3. The relationship between somatic mutations and copy number variations (CNVs) among sarcomas generated by discrete tumor initiating events. **A**, The number of genes affected by mutations versus the number of genes affected by CNVs within each sarcoma sample. **B**, The ratio of the number of genes affected by mutations to the number of genes affected by CNVs. In panels A and B, the dashed line indicates equal numbers of mutations and CNVs. **C**, The number of COSMIC genes affected by nonsynonymous mutations per tumor. **D**, The number of COSMIC genes affected by CNVs per tumor. In panels C and D, horizontal lines indicate median values for each cohort. All panels illustrate the data for n=37 tumors.
Figure 4. Nonsynonymous mutations in COSMIC genes across murine soft-tissue sarcomas. A, Mutations in COSMIC genes that occur in more than one IR-induced or Kras\textsuperscript{G12D}, p53\textsuperscript{−/−} sarcomas. B, Mutations in COSMIC genes that occur in more than 50% of MCA-induced p53 WT or MCA-induced p53\textsuperscript{−/−} sarcomas. In both panels, genes are ordered within type by the number of samples with mutations. Panels A and B illustrate the data for 19 and 18 tumors, respectively.
Figure 5. Copy number variations in COSMIC genes across murine soft-tissue sarcomas. A, COSMIC oncogenes that show a mean copy number gain and COSMIC tumor suppressor genes that show a mean copy number loss in 37 murine sarcomas. Genes are ordered within gene type by mean copy number across all samples. B, CNVs of chromosomes 6 and 9 of seven IR-induced sarcomas. DNA deletions and duplications are labeled by blue and red, respectively. Circles indicate amplicons that encompass Met on chromosome 6 and Birc3 on chromosome 9. C, Detection of Met, Birc3 and Yap1 DNA amplification in seven IR-induced sarcomas by quantitative PCR. Error bars represent mean ± SEM for three technical replicates.