

ARTICLE

<https://doi.org/10.1038/s41467-019-14261-4>

OPEN

# Mutational signatures in tumours induced by high and low energy radiation in *Trp53* deficient mice

Yun Rose Li<sup>1,2,10</sup>, Kyle D. Halliwill<sup>1,3,10</sup>, Cassandra J. Adams<sup>1,4,10</sup>, Vivek Iyer<sup>5</sup>, Laura Riva<sup>5</sup>, Rashid Mamunur<sup>5</sup>, Kuang-Yu Jen<sup>1,6</sup>, Reyno del Rosario<sup>1</sup>, Erik Fredlund<sup>1,7</sup>, Gillian Hirst<sup>1</sup>, Ludmil B. Alexandrov<sup>8</sup>, David Adams<sup>5\*</sup> & Allan Balmain<sup>1,9\*</sup>

Ionising radiation (IR) is a recognised carcinogen responsible for cancer development in patients previously treated using radiotherapy, and in individuals exposed as a result of accidents at nuclear energy plants. However, the mutational signatures induced by distinct types and doses of radiation are unknown. Here, we analyse the genetic architecture of mammary tumours, lymphomas and sarcomas induced by high (<sup>56</sup>Fe-ions) or low (gamma) energy radiation in mice carrying *Trp53* loss of function alleles. In mammary tumours, high-energy radiation is associated with induction of focal structural variants, leading to genomic instability and *Met* amplification. Gamma-radiation is linked to large-scale structural variants and a point mutation signature associated with oxidative stress. The genomic architecture of carcinomas, sarcomas and lymphomas arising in the same animals are significantly different. Our study illustrates the complex interactions between radiation quality, germline *Trp53* deficiency and tissue/cell of origin in shaping the genomic landscape of IR-induced tumours.

<sup>1</sup>UCSF Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA 94158, USA. <sup>2</sup>Department of Radiation Oncology, University of California San Francisco, San Francisco, CA 94143, USA. <sup>3</sup>Abbvie, Redwood City, CA 94063, USA. <sup>4</sup>Nuffield Department of Medicine, University of Oxford, Oxford OX7DQ, UK. <sup>5</sup>Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. <sup>6</sup>Department of Pathology, University of California Davis Medical Center, Sacramento, CA, USA. <sup>7</sup>Doublestrand Bioinformatics, 11331 Stockholm, Sweden. <sup>8</sup>Department of Cellular and Molecular Medicine and Department of Bioengineering, Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA. <sup>9</sup>Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94158, USA. <sup>10</sup>These authors contributed equally: Yun Rose Li, Kyle Halliwill, Cassandra Adams. \*email: [da1@sanger.ac.uk](mailto:da1@sanger.ac.uk); [allan.balmain@ucsf.edu](mailto:allan.balmain@ucsf.edu)

Carcinogens in the environment contribute to cancer development by inducing DNA damage, resulting in specific patterns of mutations that can be diagnostic for the causative carcinogenic agent. Such “mutational signatures” help us to understand and quantify the impact of potential carcinogens on human tumour DNA, and can provide information on the mechanisms by which these agents act. Mutational signatures have been detected in thousands of human cancers<sup>1</sup>, as well as in mouse models of chemical carcinogen-induced tumours<sup>2,3</sup>. In contrast, while ionizing radiation through medical exposures has been associated with increased risk of secondary malignancies<sup>4</sup>, we presently have no detailed knowledge of the types of signatures that are induced by this common source of DNA damage.

Environmental and therapeutic IR can exist in a number of different forms, depending on radiation quality and energy, which impact how radiation interacts with biological target molecules, such as DNA. High energy gamma rays or photons are the most commonly used form of therapeutic IR and occur naturally in the earth’s atmosphere due to radioactive decay or are secondary to the action of cosmic rays. In contrast, heavy-ion or high linear energy transfer (LET) radiation is characterised by charged, high energy particles (protons, Fe-ions) and accounts for approximately 1% of cosmic rays<sup>5,6</sup>. Individuals at highest risk from radiation exposures are patients being treated for cancer by localised radiotherapy, most often using fractionated high doses of gamma rays to focally targeted sites. However lower doses of radiation also have biological effects, the consequences of which have not been adequately investigated in terms of possible genomic changes.

In fact, the exact mechanisms through which radiation induces DNA damage and its impact on the genomic landscape of radiation-induced malignancies are not well-established and have been highly controversial. Although there is general support for a direct model in which IR causes single or double strand breaks by direct collision with DNA, indirect models have also been proposed by which IR interacts with other molecules (e.g., water) to increase DNA damage secondary to free radical formation, or impacts the normal tissue microenvironment to promote cytokine production and transformation through mechanisms that are presently unclear<sup>7</sup>. Radiation has been proposed to cause transformation due to a “bystander effect” even in cells that have not been directly traversed by radiation particles<sup>8,9</sup> but the role of this mechanism in vivo has not been established. Resolution of these questions is important not only for our understanding of mechanisms of carcinogenesis after therapeutic or accidental exposure to ionizing radiation, but also for assessment of cancer risk as a result of occupational exposure, or, increasingly, as a consequence of longer term exposure to cosmic radiation during space flights<sup>10,11</sup>.

Characterising these mechanisms in human tumours is impeded by limited sample size, germline heterogeneity, complex environmental exposures and uncertainty in the dose, duration, frequency and quality of radiation exposure. Thus, the use of appropriate model systems that properly recapitulate the effects of RT in cells and animal models is pivotal. Indeed, high and low LET studies have been carried out on cells in vitro and in lower organisms, and the types of changes seen depend on radiation quality (reviewed in Durante & Cucinotta)<sup>12</sup>. Specifically, high LET radiation is associated with intense focal damage caused by the traversal of heavy ions through the nucleus<sup>13–15</sup>. In contrast, gamma radiation causes more indirect DNA damage, diffuse patterns of H2AX foci, and increased oxidative stress<sup>16</sup>.

*Trp53*, a gene also known as the guardian of the genome<sup>17</sup>, is critical in affecting the cellular outcome following DNA damage<sup>18,19</sup>. Human patients with germline *TP53* mutations

are highly tumour-prone<sup>20,21</sup>, and genomic studies have linked germline *TP53* deficiency to development of genomically unstable tumours with a high incidence of chromothripsis<sup>22–25</sup>. To develop a deeper understanding of the effects of radiation on tumour DNA integrity, we use mouse strains with partial germline deficiency in *Trp53* function, i.e., mice hemizygous for *Trp53* (*Trp53*<sup>+/-</sup>), or homozygous for the Delta P *Trp53* allele (*Trp53*<sup>(DP/DP)</sup>); hereafter *Trp53ΔP*). The *Trp53*<sup>+/-</sup> strain is highly susceptible to both spontaneous and radiation-induced (4 Gy gamma) lymphoma development, while the *Trp53ΔP* strain develops a wider range of tumours after exposure to the same radiation dose<sup>26–28</sup>. The *Trp53ΔP* mouse carries a deletion in the N-terminal proline-rich domain (PRD) of TP53, a region that is important in the regulation of human TP53 activity and stability. A common germline polymorphism at codon 72 in this region has been associated with altered cancer risk in humans<sup>29,30</sup>.

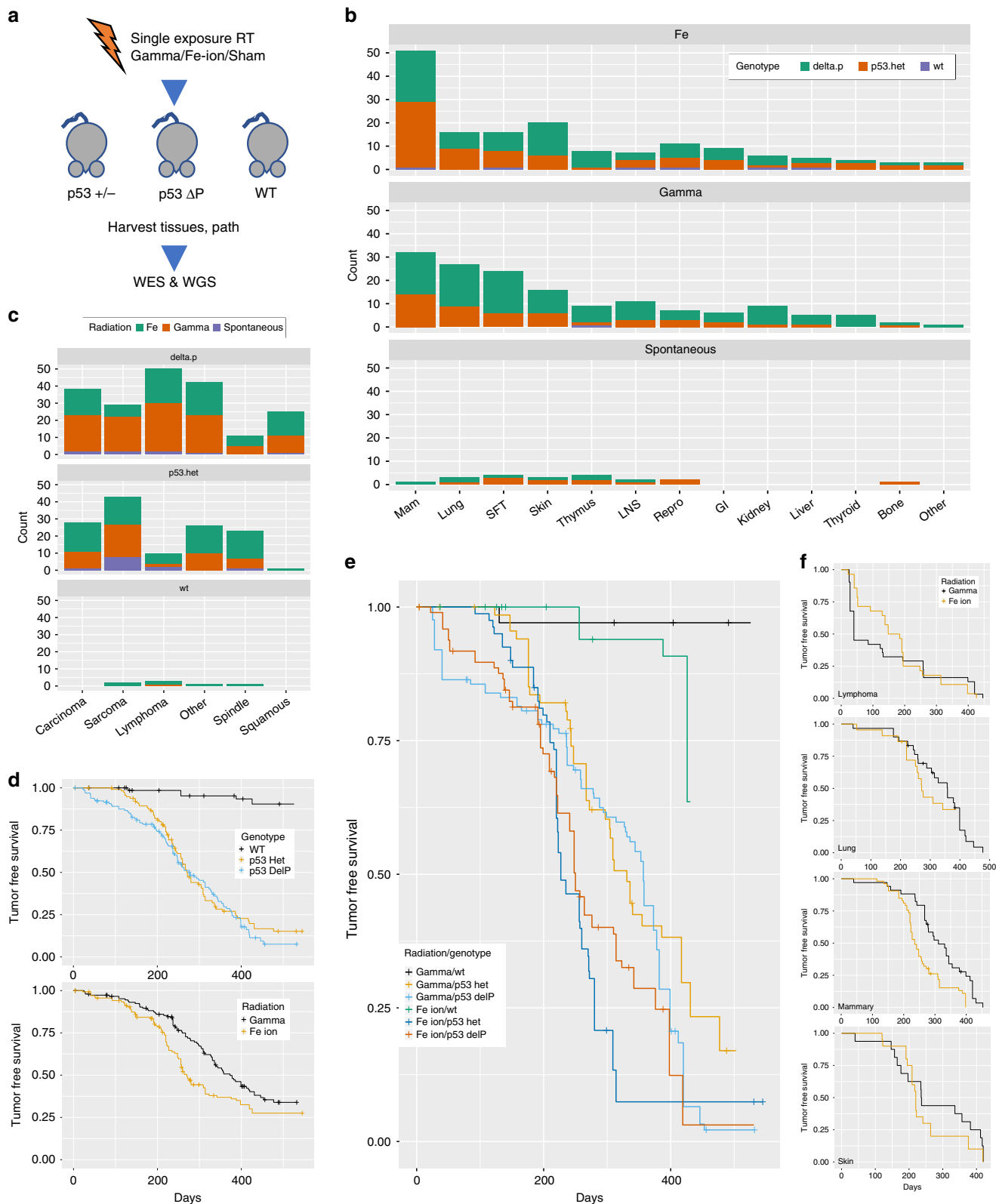
In this study, we breed both strains of mice on to the same genetic background (FVB/N), and expose the animals to a relatively low dose (50 cGy) of gamma (low LET) or Fe-ion (high LET) radiation to explore the effects on susceptibility to tumour development in different tissues. In particular, the 50 cGy dose of Fe-ions is studied as it has been proposed by NASA and other previous studies as a realistic dose to which astronauts may be exposed during exploratory space missions<sup>31–33</sup>. We demonstrate that this low radiation dose results in the development of a wide range of malignancies in the context of germline *Trp53* deficiency, but particularly in a high incidence of mammary carcinomas and angiosarcomas, which are also common radiation-associated malignancies in humans. Whole-genome sequencing of mammary tumours and whole-exome sequencing of a wider spectrum of IR-induced tumours reveals complex point mutational signatures and patterns of structural rearrangements in tumour genomes that are associated with radiation quality, as well as germline *Trp53* status and tumour histology.

## Results

### Low dose radiation-induced tumours in *Trp53*-deficient mice.

We previously showed that *Trp53ΔP* mice in a mixed 129/C57BL6 background were highly sensitive to development of several tumour types after a single exposure to whole body gamma radiation, but had a low spontaneous rate of tumour development<sup>26</sup>. We exposed 238 FVB/N mice that were *Trp53* WT, hemizygous (*Trp53*<sup>+/-</sup>) or homozygous for the Delta P *Trp53* allele (*Trp53ΔP*), to a single dose of either <sup>56</sup>Fe-ion (600 MeV/amu) or gamma radiation. Most animals were exposed to a dose of 50 cGy of either radiation type, but a few received lower or higher doses (Supplementary Data 1 and Supplementary Table 1). Animals were followed for >400 days to monitor tumour development; a subset of sham irradiated controls was also included (Fig. 1a). A total of 160 mice developed at least one of several histological types of tumour after radiation exposure, most commonly in the mammary glands, skin, lung and lymphoid tissue (Fig. 1b, c). As expected, wild type mice were generally resistant to development of tumours induced by radiation (Fig. 1d), but a few animals were sacrificed early, predominantly those that developed tumours after exposure to <sup>56</sup>Fe-ions (Fig. 1e). Overall, *Trp53ΔP* animals were much more likely to develop carcinomas in the skin, mammary gland, and other tissues, versus other histologic subtypes, ( $P = 3.63 \times 10^{-5}$ , chi-sq test; Fig. 1b, c).

The most common site of tumour development was the mammary gland, regardless of radiation quality or *Trp53* germline status (Fig. 1b). There were a number of histological subtypes, including adenocarcinomas (ADC), squamous cell carcinomas (SCC), spindle cell carcinomas (SPN), and sarcomas (SAR), including angiosarcomas (Fig. 1c). Notably, in humans, sarcomas



and angiosarcomas appear to show a disproportionately higher incidence in patients with a history of prior radiation therapy to the chest or thorax<sup>34</sup>. Compared to other models of radiation-induced tumours in *Trp53*-deficient mice (129/Sv strain, 4 Gy gamma ray exposure) where the predominant tumour types were lymphomas and sarcomas<sup>27,35</sup>, the present model (i.e., FVB/50

cGy) more closely mimics the spectrum of radiation-induced tumours observed in humans.

Radiation quality had a significant impact on tumour latency and disease-free survival (DFS), as animals exposed to Fe-ion radiation survived on average 287 days, as compared to gamma radiation (358 days) (univariate Cox-Proportional Hazard analysis,

**Fig. 1 Tumour histology and survival in observed RT-induced malignancies.** **a** Study schema showing that animals of different germline *Trp53* status (*Trp53*<sup>+/-</sup>,  $\Delta P$  (*Trp53\Delta P*), and WT) receiving radiation of different qualities (Gamma, Fe-Ion and Sham) developed a range of different tumour types that were harvested and studied for pathology and selectively sequenced by whole-exome (WES) or whole-genome (WGS) platforms. **b** Tissue distribution of different tumour types following whole body radiation exposure, as a function of radiation quality and germline *Trp53* status. Independent tumours from animals (with clearly independent primaries by histology) were separately counted. SFT soft tissues, LNS Lymph nodes, Repro, reproductive organs, GI gastrointestinal tract. *N* = 479. **c** Incidence of different pathologic tumour types by radiation quality and germline *Trp53* status. Most adenocarcinomas and sarcomas were in the mammary gland, and most squamous carcinomas in the skin. *N* = 479. **d** Tumour-free survival (TFS) of animals by germline genotype (top) and radiation quality (bottom). An event was defined death due to tumour confirmed by pathology and dissection and/or visible tumour requiring euthanasia. *N* = 238. **e** Interaction of genotype and radiation quality on TFS. Wildtype animals with exposure to Fe-Ion have a substantial risk of malignancy as compared to animals with gamma exposure. *N* = 238. **f** TFS related to different tumour disease sites. Survival was reduced in animals developing skin, mammary or lung tumours after exposure to Fe ions compared to gamma radiation, but this was reversed for lymphomas, which showed earlier onset and were preferentially induced by gamma radiation exposure. *N* = 238.

HR = 2.36,  $P < 1.94 \times 10^{-6}$ ; Fig. 1d–f). No obvious effect on survival was attributable to mouse strain genotype (*Trp53*<sup>+/-</sup>) or *Trp53\Delta P*, Fig. 1d). On analysis by disease site, mammary, skin, and lung tumours all had a shorter latency following exposure to <sup>56</sup>Fe radiation compared to gamma. In contrast, a different pattern was seen for lymphomas, which developed earlier after exposure to gamma radiation (Fig. 1f, lower right panel).

**Mutational signatures in radiation-induced tumours.** We performed whole-genome sequencing (WGS) of a focused cohort of 21 mammary tumours from animals of *Trp53*<sup>+/-</sup> or *Trp53\Delta P* backgrounds exposed to 50 cGy Fe-ion or gamma radiation (Supplementary Table 1). The overall frequency of point mutations or single nucleotide variants (SNVs) was low (median = 1.16, mean = 2.009; range 0.62–14.6 SNVs/Mb), considerably less than observed in chemically-induced mouse tumours of the same genetic background<sup>2,3</sup>. The rate did not differ significantly according to tumour type or radiation quality ( $P = 0.52$  and  $P = 0.92$ , respectively, Fisher's Exact test; Supplementary Fig. 1A). The dearth of point mutations in these samples, most of which mapped to non-coding regions of the genome (Supplementary Fig. 1C and D), is consistent with that reported previously in radiation-associated human tumours<sup>4</sup>.

As SNV mutational signatures can provide genomic evidence for the aetiology of tumour initiation and progression, we next evaluated whether tumours that result from exposures to different radiation qualities show distinct point mutation patterns<sup>1,36,37</sup>. We hypothesised that a signature of reactive oxygen species (ROS)-related DNA damage may be observed as an indirect consequence of IR exposure. IR is known to generate free radicals<sup>38</sup>, which can in turn damage DNA by causing the formation of 8-Oxo-guanine adducts. We addressed this question by two different approaches: searching for de novo mutational signatures using Nonnegative Matrix Factorisation (NMF), and analysis of cosine similarity to existing trinucleotide mutational signatures in a wide range of human cancers<sup>37</sup>. When we performed de novo mutational signature extraction using WGS-derived SNVs, we observed 3 independent signatures (Fig. 2a). Signature SNV A was enriched for G > T (C > A) substitutions which can be generated by misrepair of 8-oxo-guanine adducts<sup>38</sup>. This signature is consistent with the previously described human COSMIC signature 18, which was previously reported to be associated with oxidative DNA damage due to ROS<sup>39,40</sup>. Signature SNV B resembles the pattern of human signature 5, which is one of the clock signatures and is comprised of C > T and T > C transitions<sup>37</sup>, while signature SNV C had no obvious similarity with other Cosmic signatures.

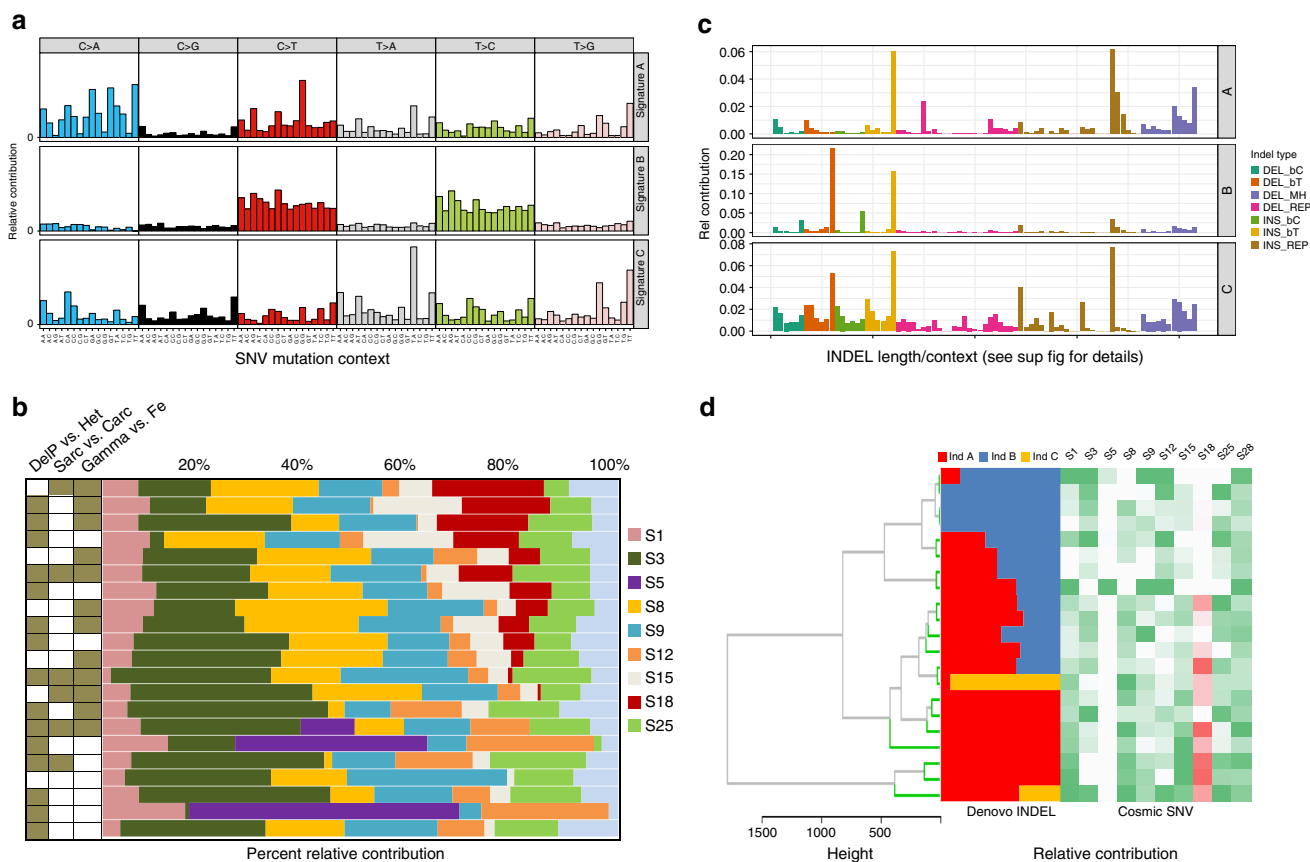
Signature deconvolution using the established human COSMIC mutational signatures<sup>36,37</sup>, revealed a strong cosine similarity with several human signatures, including signature 18, clock signature 5, and signature 3, which is found in most mouse mammary

carcinomas and is associated with defects in homologous recombination (Fig. 2b). This process is critical for the repair of double-stranded breaks induced by RT<sup>37</sup>. Signature 18 (red, Fig. 2b) was variable across samples, but was preferentially enriched in tumours induced by gamma radiation, compared to those induced by <sup>56</sup>Fe-ions ( $p < 0.032$ , Fisher's Exact test). While these data are based on a relatively small sample size (12 gamma-induced and 9 heavy ion-induced tumours), the results are consistent with the hypothesis that 50 cGy gamma radiation results in higher levels of ROS than 50 cGy high LET radiation.

**Indel signatures in radiation-induced tumours.** We performed de novo INDEL mutational signature extraction from the 21 WGS samples, identifying three unique INDEL signatures: A, B and C (Fig. 2c and Supplementary Fig. 2)<sup>1</sup>. Signature INDEL-A was enriched for microhomology domain-associated deletions and repeat-mediated insertions, while signature INDEL-B was enriched in single base simple Thymidine insertions and deletions. Both signatures were observed in the majority of samples without clear association with *Trp53* genotype. Signature INDEL-C was identified in only 2 samples and was not strongly enriched for any particular INDEL class.

Unsupervised hierarchical clustering of both INDEL and SNV signatures (Fig. 2d) revealed some significant correlations between specific combinations of INDELS and SNVs. The high ROS SNV signature (human signature 18) that was enriched in samples induced by exposure to gamma radiation (Fig. 2b) showed a significant correlation with signature INDEL A ( $P = 0.017$ , Fisher's Exact test) (Fig. 2d). Interestingly, Sharma et al. have proposed that ROS-induced damage can cause oxidatively induced clustered DNA lesions that lead to double strand breaks and deletions caused by error-prone repair by the non-homologous end-joining (NHEJ) pathway<sup>41</sup>. It is therefore possible that the co-occurrence of SNV signature 18 and signature INDEL A is related to relatively higher levels of ROS induced by gamma rays, but further detailed studies would be required to confirm this association.

**Structural rearrangements in radiation-induced tumour genomes.** In contrast to the low SNV rate, genomes from radiation-induced tumours bore frequent structural rearrangements. Figure 3 shows the patterns of genomic alterations in mammary tumours arranged by sample ID (Fig. 3a) or by chromosome (Fig. 3b) based on whole-genome sequencing. We identified a mean of 27 rearrangements per tumour. These large structural variants (SVs) were enriched in deletions (mean  $n = 11.3$ /sample) as opposed to insertions (1.52) or inversions (6.1). A total of 74 intrachromosomal (ITX) and 66 interchromosomal complex translocations (CTX) were observed. The distributions of mutations across samples and chromosomes were non-homogeneous.



**Fig. 2** SNV and INDEL signatures in RT-induced mammary tumours. **a** De novo SNV signature extraction using NMF identified three distinct point mutation signatures, referenced to here and in the text as “A”, “B” and “C”.  $N = 21$ . **b** Deconvolution of the mutation signatures into known Cosmic signatures found in human tumours. Samples are denoted by radiation quality, tumour pathology and *Trp53* genotype. DelP (Brown) = P53 Delta Proline (*Trp53ΔP*); Het = P53 Heterozygous (White), Sarcoma (Brown), Carcinoma (White), Gamma (Brown) and Fe (White). The red blocks represent the contribution of the signature 18 “ROS” to total mutations in each sample, and these are relatively enriched in tumours induced by gamma radiation compared to  $^{56}\text{Fe}$  ions ( $P < 0.032$ ). Pale blue denotes contribution from variants not explained by known Cosmic signatures.  $N = 21$ . **c** Relative contribution of each insertion/deletion mutation type in the three INDEL signatures identified in the mammary tumour samples. See Supplementary Fig. 2 for a breakdown of all categories of INDEL types. “Del\_B” = 1 bp deletion, “Ins\_B” = 1 bp insertion, where “C” or “T” describes whether cytosine or thymidine was deleted at that position, “MH” = microhomology domain and “REP” = >1 bp deletion or insertion at repeats.  $N = 21$ . **d** Relationship between INDEL signatures and SNV signatures. Hierarchical clustering of both INDEL signatures and SNV signatures demonstrates an enrichment of de novo INDEL signature A (left) and SNV signature 18 (ROS, right) in samples induced by exposure to gamma radiation compared to those induced by  $^{56}\text{Fe}$  ions ( $P < 0.032$ ).  $N = 21$ .

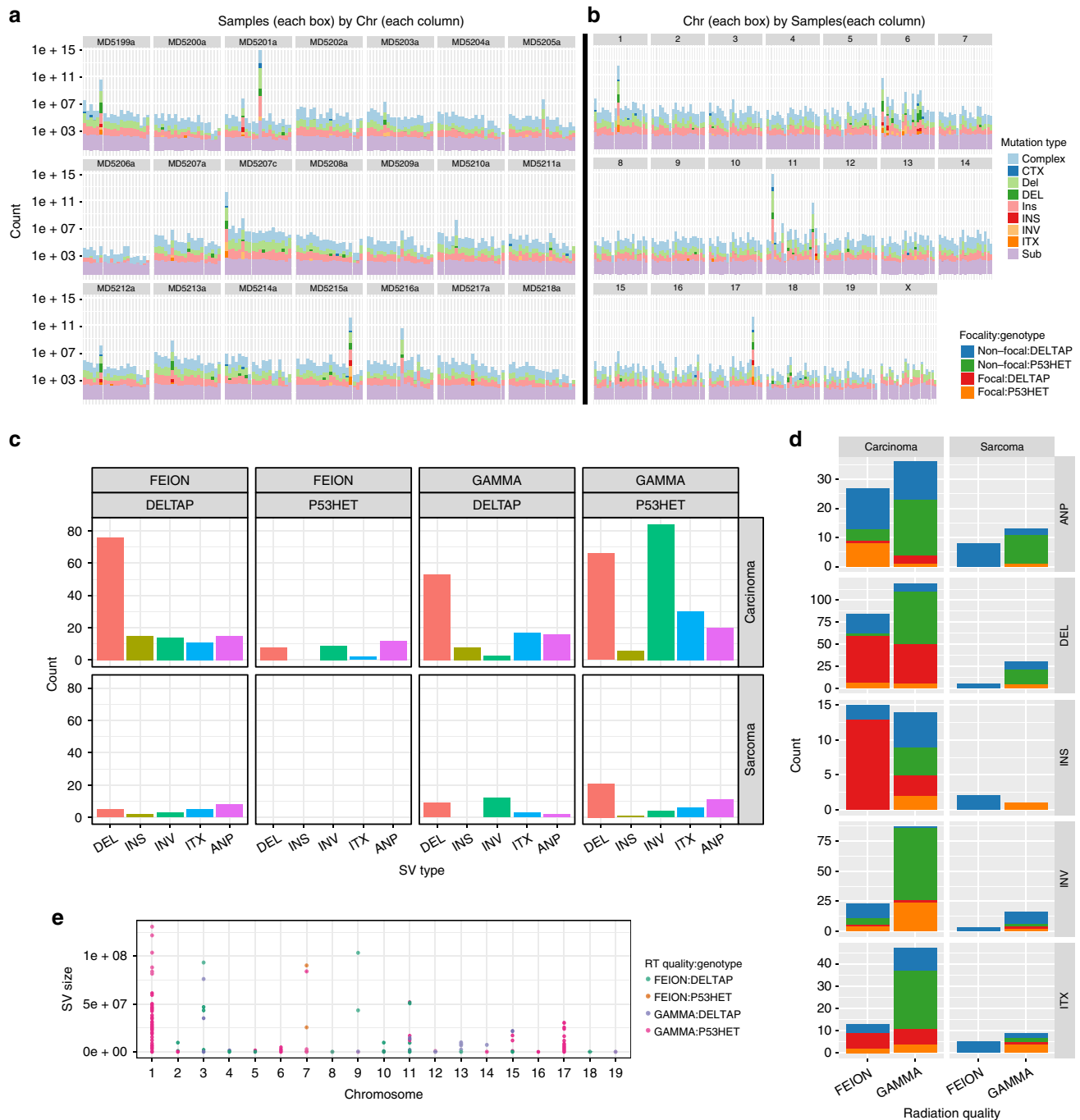
SVs as well as smaller INDELS were clustered in some tumours (Fig. 3a) and also were located predominantly in a few chromosomes including chromosomes 1, 6, 11 and 17 (Fig. 3b).

Carcinomas were significantly more enriched in large scale SVs than sarcomas ( $P < 0.01$ , Fisher’s Exact test; Fig. 3c top vs lower panels). Furthermore, inversion rearrangements were more common in carcinomas from *Trp53+/-* than from *Trp53ΔP* mice, and in particular carcinomas in mice exposed to gamma radiation carried 68.2% of all inversions ( $P < 1 \times 10^{-4}$ , Fisher’s Exact test; Fig. 3c (top right panel) and 3D). Prior work has shown that distinct mutational processes can impact the observed frequency of focal versus non-focal rearrangements<sup>42</sup>. We classified structural variants into focal and non-focal SV types and examined the differences in the distribution of focal SVs (including mixed chromosomes containing complex regions of copy number gains and losses) as opposed to non-focal rearrangements (including whole chromosome aneuploidy (see Methods section), as a function of germline *Trp53* genotype, histological subtype and radiation quality (Fig. 3d). In addition to showing that SVs are rare in sarcomas, this analysis showed that  $^{56}\text{Fe}$ -ion exposure was more likely to result in focal deletions and insertions, whereas non-focal

SVs or whole chromosome aneuploidy were enriched in tumours induced by gamma radiation or from *Trp53+/-* mice (Fig. 3d). Most SVs were small (<100 Kb) although clusters of larger SVs were observed in several chromosomes (Fig. 3e). We conclude that the type of radiation to which the mice are exposed has differential effects on the nature of the resulting genomic alterations, with heavy ion, high LET radiation linked to more focal changes, particularly in *Trp53ΔP* mice, and low LET gamma radiation associated with larger scale chromosome alterations, particularly in *Trp53* heterozygous mice. The data therefore demonstrate a strong interaction between germline *Trp53* status and radiation quality in determining patterns of genomic instability.

**Gene copy number changes in radiation-induced tumours.**

Analysis of gene copy number variation (CNVs) in the WGS data revealed striking patterns of gains or losses across many different chromosomes (Supplementary Fig. 3). Copy number gains were seen frequently on chromosomes 5, 6 and 15, while losses predominated on chromosomes 4, 9, 12 and 18. Interestingly, some individual tumours showed patterns dominated mainly by CNV



**Fig. 3 The mutation landscape of SNVs and SVs across radiation-induced malignancies.** **a** Distribution of mutations identified in mammary tumours by individual samples. Each column is one chromosome (*x*-axis). Total count of mutations is on the *y*-axis. Mutations are coloured by type. Sub: Substitutions (SNVs), ITX: Intrachromosomal translocation, INV: Inversion rearrangements, INS: insertion rearrangements, DEL: deletion rearrangements, CTX: Complex or inter-chromosomal translocations, Complex: Complex insertions. Number of base pairs impacted by each mutation type is plotted on the *y*-axis in logarithmic scale. **b** Distribution of mutations (as in **a**) in each WGS mammary tumour sample sorted by chromosome. Each column corresponds to one unique sample and each facet box encompasses one chromosome. Large scale variants were found mainly in chromosomes 1, 6, 11 and 17. **c** Frequency (*x*-axis) of structural variants observed in mammary tumours based on radiation quality, *Trp53* genotype and tumour histology. ITX: intrachromosomal translocation, INV: inversion rearrangements, INS: insertion rearrangements, DEL: deletion rearrangements, ANP: aneuploidy. Carcinomas have more SVs than sarcomas from the same cohort of mice, and large inversions are enriched in tumours from *Trp53*<sup>+/-</sup> mice exposed to gamma radiation. **d** Incidence of focal and non-focal genomic events in mammary tumours based on radiation quality, germline *Trp53* genotype and tumour histology. A higher rate of focal mutations (red and orange blocks) is observed in Fe-Ion associated carcinomas. Large non-focal structural variants (green and blue blocks) were more prevalent in samples induced by gamma radiation. Mammary sarcomas have significantly fewer structural variants than mammary carcinomas from the same cohorts of mice, independent of radiation quality and *Trp53* genotype. **e** Distribution of structural variants identified in the same samples as in **c** by size across the mouse genome. Size (*y*-axis) is shown by log scale. Chromosome number is shown along the bottom. Most SVs are small and less than 1 MB, as expected.



**Fig. 4 The landscape of CNVs illustrate both focal and whole-chromosome events.** **a** CNVs across all sequenced (whole exome and whole genome) samples. Tumour samples are sorted by tissue site (large blocks separated by black lines) and subsequently by pathology, radiation quality and genotype status of *Trp53*. Relative copy number states across each chromosome are indicated by a corresponding key as shown at the lower right. Some chromosomes show predilection for whole-chromosomal amplification (e.g. Chromosome 5 and 15, the latter particularly in lymphomas) while other regions show more deletions (Chromosome 4) or focal amplifications (e.g., proximal Chromosome 6 corresponding to *Met* in mammary samples). **b** Circos plot illustrating relationships between amplification status, gene expression, translocations, and candidate driver genes across tumours induced by radiation exposure. From outside to inside: 1. Driver genes. 2. Cytoband map of the mouse chromosome. 3. Focal CNVs gains. 4. Focal CNVs losses. 5. Aneuploidy gains. 6. Mixed chromosome changes. 7. Aneuploidy losses. 8. Expression of cancer driver genes in tumour samples versus normal samples. 9. Intrachromosomal and interchromosomal translocations identified in samples that were subjected to whole-genome sequencing. Y-chromosome was not included in the analysis.

*Trp53* genotype or radiation quality. Chromosome 15 aneuploidy has previously been found to be a common genetic event in gamma radiation-induced lymphomas, and is thought to be driven by *cMyc*<sup>35,43</sup>. These data demonstrate the tissue specificity of genomic changes in tumours of distinct tissue and cellular origins, even in the same individual animal and after exposure to the same carcinogenic stimulus. While CNV analysis (Fig. 4) showed numerous examples of whole chromosome aneuploidy, defined as a contiguous region of gene copy number gains or losses (see Methods section), other samples had chromosomes with a mixed pattern comprising both gains and losses. The latter phenomenon was significantly correlated with radiation quality as 85% of mixed chromosomes were associated with Fe-ion exposure ( $P < 0.03$ ). This is also consistent with the data above showing that <sup>56</sup>Fe-ion exposure is associated with increased focal deletions and insertions, supporting a model in which <sup>56</sup>Fe-ion radiation causes direct focal damage in DNA, leading to double strand breaks (DSB). There was also an association with tumour type, as mammary tumours were strongly deletion biased (52%) as compared to lymphoma (21%) or skin tumours (15%) ( $P < 2.9 \times 10^{-3}$ , chi-sq test).

Aside from chromosome 15 gains, the vast majority of aneuploid chromosomes were observed in mammary tumours. For example, loss of chromosome 4 was observed in 14 samples, 11 of which were mammary tumours. Chromosome 4 contains *Cdkn2a* encoding the *Ink4/Arf* locus strongly implicated in cancer pathogenesis with well-studied functional links to *Trp53*. Focal loss of *Cdkn2a* is a common event in chemically induced mouse tumours<sup>3</sup>, but such focal events were rarely detected in radiation-induced tumours. This may suggest that other genes on chromosome 4 are also driving the observed losses.

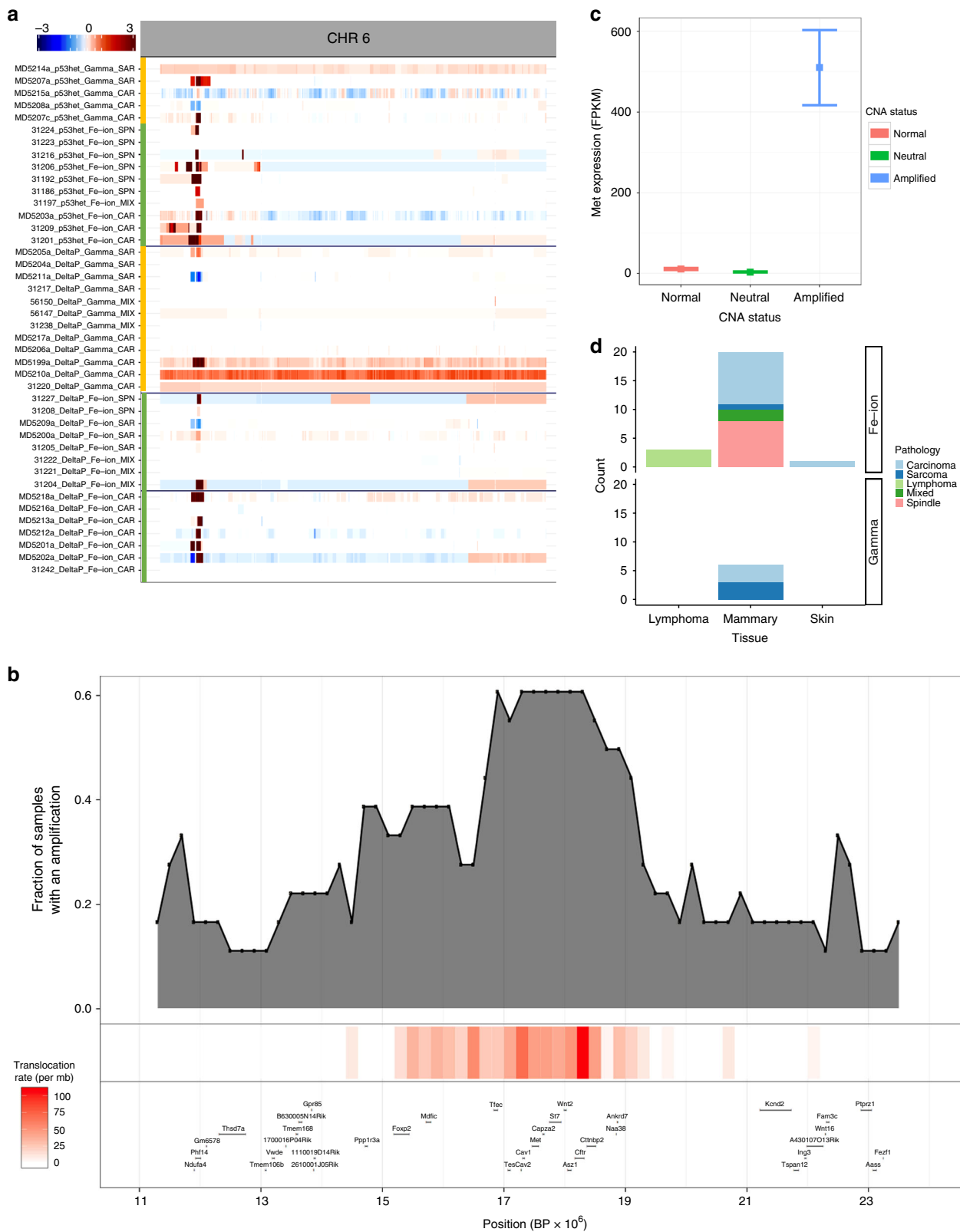
**Met is a common target of amplification in heavy ion-induced mammary carcinomas.** The proximal region of chromosome 6 encompasses the *Met* oncogene, which was amplified in a total of 23 mammary tumour samples, 19 of which were induced by <sup>56</sup>Fe-ions and only 4 by gamma radiation (Fig. 5a, d). These data suggested that the focal genomic events preferentially induced by heavy ions select for mammary tumours driven by *Met* amplification. RNASeq analysis of a subset of mammary tumours confirmed that amplification of the *Met* locus was indeed associated with a large increase (>200 fold) in *Met* expression (Fig. 5c). A functional contribution from other genes in the amplified region which were also over-expressed, although not to the same level (Supplementary Fig. 4A) is however also possible. The rearrangement hotspot on chromosome six overlapped the significantly amplified region chromosome 6 cytoband A2, however the peak region of structural rearrangements did not exactly overlap with the *Met* gene (Fig. 5b). This is compatible with a model by which complex structural SVs can lead to progressive rearrangements and selection for increased copy number and expression of a strong cancer driver such as the *Met* gene<sup>24</sup>.

**Germline *Trp53*-dependent patterns of structural variants.** Chromothripsis, or chromosome-shattering, is a phenomenon by which numerous clustered chromosomal rearrangements occur in localised, confined genomic regions, first characterised in human tumours and some congenital diseases. This process has been attributed to a single catastrophic genetic event that results in extensive DNA damage in the context of *Trp53* dysfunction<sup>24,25,44–46</sup>. Chromothripsis is particularly associated with tumour development in patients with Li-Fraumeni syndrome as they bear germline *TP53* mutations<sup>22</sup>, but may also be more generally associated with partial insufficiency in *TP53* or other genes linked to DNA damage responses such as *ATM*<sup>22,47</sup>. While SVs were the predominant lesions observed in mouse radiation-associated tumours, we also saw examples of whole or partial chromosome chromothripsis. One mammary tumour (MD5207c) showed strong evidence for whole chromosome chromothripsis, with clusters of SVs across chromosome 1 (Fig. 6a, b). This mammary carcinoma was from a gamma-exposed *Trp53*+/- mouse, and showed many of the hallmarks of chromothripsis including chromosome wide translocations, deletions and inversions, with several showing overlapping breakpoints consistent with a simultaneous rather than sequential event<sup>24</sup>.

Some carcinomas showed complex deletions on chromosome 11 overlapping *Trp53* (Fig. 7a). In total, at least five carcinomas showed focal clustered regions of SVs on chromosome 11, three of which had variants clearly disruptive of the *Trp53* locus. The two other samples demonstrated focal clustering of SVs on chromosome 11 which mapped close to but distal from *Trp53* (Supplementary Fig. 5). Notably, CNV analysis of chromosome 11 identified distinct patterns of changes in tumours from *Trp53*+/- and *Trp53* $\Delta P$  mice (Fig. 7b, top panel vs lower panels). Tumours from *Trp53*+/- mice showed a higher frequency of large scale or whole chromosome loss of chromosome 11, regardless of radiation quality, as a common mechanism for ensuring complete loss of *Trp53* function (Fig. 7b). In contrast, almost all changes of copy number at the *Trp53* gene in tumours from *Trp53* $\Delta P$  mice were seen in <sup>56</sup>Fe-ion radiation-exposed animals (Fig. 7c), suggesting that focal DNA damage of the type induced by heavy ions is more likely to result in loss of *Trp53* in mice with this mutant allele, which is known to retain some residual *Trp53* function<sup>28</sup>.

**Common driver events in radiation-induced tumours.** Several other potential driver genes in radiation-induced tumours were identified by a combination of CNV analysis, exome sequencing to identify point mutations, and gene expression (Supplementary Data 3). One amplicon seen in several samples was on proximal chromosome 9 in a region harbouring a family of genes encoding matrix metalloproteinases (Mmps). MMPs can regulate the tumour microenvironment, and their expression and activation is increased in tumour tissue compared to normal<sup>48</sup>. Specific *Mmp* genes, including *Mmp* 13, 12, 10 and 3, were highly expressed in





tumours carrying the amplified region (Supplementary Fig. 4B). However alternative candidates *Birc2*, *Birc3* and *Yap1*, also localised in the amplified region, showed some level of over-expression and may also be functionally significant.

We also observed recurrent gene-level events in *Notch1*, a gene known to play an important role in radiation-induced lymphoma

development<sup>49</sup>. In total, five lymphoma samples had distinct *Notch1* mutations. In two lymphomas, we found frameshift indels in the PEST domain, resulting in activation of Notch signalling, as well as multiple nonsynonymous SNVs in *Notch1* (Supplementary Data 3 and 4). Unlike the indels, these were not limited to the PEST domain and their functional implications are unclear. In

**Fig. 5 Focal amplification of *Met* is preferentially induced by  $^{56}\text{Fe}$ -ion radiation.** **a** Mammary tumours bear high frequency of focal amplifications on chr 6 (at the *Met* locus) predominantly in Fe-ion associated tumours.  $N = 42$ . **b** Focal analysis of the *Met* locus to illustrate that the same region with focal *Met* amplifications are co-enriched for structural variants. The highest frequency of structural variants (red vertical line) does not coincide exactly with the locus with the increased copy number of *Met*, suggesting that other selection mechanisms led to the high *Met* copy number. Locations of genes in the amplified region are shown in the lower panel.  $N = 21$  (translocations from WGS data).  $N = 10$  (expression from RNAseq data). **c** Higher expression of *Met* by whole transcriptome RNA-seq is strongly associated with copy number status. RNA from a total of 10 mammary tumours was sequenced. Expression is shown for tumours with or without *Met* amplification, as well as matched normal samples. Error bars = 95% confidence intervals. **d** Association between *Met* amplification and tumour site of origin (mammary, skin, lymphoma), histology and radiation quality.

two gamma radiation-induced lymphomas we found focal amplification of *Notch1*, one of which also carried a point mutation, demonstrating the multiplicity of mechanisms leading to activation of *Notch1* in radiation-induced lymphomas. Finally, significant SNVs were found in a range of known driver genes including *Kras*, *Apc*, *Pten* and hotspot sites in *Trp53*, the latter examples occurring in tumours from the *Trp53 $\Delta$ P* mice (Supplementary Fig. 6 and Supplementary Data 3). The spectrum of genomic alterations, both focal and large-scale, found in these radiation-induced tumour samples is complex and reflects the heterogeneity of genomic consequences resulting from radiation exposure. The changes found over all samples is summarised in the Circos plot in Fig. 4b.

## Discussion

We have explored the impact of both high and low energy radiation on tumour development in mice carrying two distinct types of germline mutation in the *Trp53* gene. Our data demonstrate that tumours induced by both high and low LET radiation exhibit the common features of low SNV burden, significant genomic instability as shown by detection of large-scale structural variants, and evidence of whole or partial chromosomal chromothripsis. However, radiation quality had a significant impact, together with germline *Trp53* status, on the number and types of tumours that developed, the overall survival of affected mice, and the classes of genomic aberrations found by whole genome and exome sequencing. The low SNV burden and high indel/SNV ratio in mouse mammary tumours mirrors results obtained by WGS analysis of a set of 12 human tumours that were associated with prior exposure to radiotherapy<sup>4</sup>, but contrasts with the results obtained by analysis of tumours induced in NF1-deficient mice by high dose fractionated ionising radiation<sup>50,51</sup>. In the latter studies based on exome but not whole genome sequencing, a much higher SNV burden was found, including many recurrently mutated genes. However these data are difficult to interpret because of the differences in the mouse models used, the experimental design and methods of analysis, which did not include filtering of SNVs through exactly matched normal control DNA samples (see also Chang, H.-H. et al.<sup>52</sup>).

A surprising feature of our experiments, which evaluated the effects of 50 cGy high or low LET radiation, was the exceptionally high incidence of radiation-induced mammary cancers, comprising carcinomas and a range of sarcoma types including angiosarcomas. Other models of radiation-induced tumour development have not generally resulted in a high incidence of mammary cancers. This could be due to the effects of higher radiation dose (in the 2–5 Gy range) which is associated with earlier onset haematologic malignancies and lymphomas. Alternatively, it could be due to a combination of dose and genetic background, which could result in early death due to aggressive lymphoma or sarcoma development before the time when secondary carcinomas would arise, as the median tumour free survival was 267 and 327 days for mammary and lung tumours while it was only 130 days for lymphomas. This effect is further

accentuated in gamma-IR (40 day latency in lymphoma versus 331 days in mammary), which has been more frequently used in historical studies. The present model therefore more closely mimics susceptibility to human secondary breast cancer development, which is known to be increased following radiation exposure and to be influenced by both germline TP53 and somatically acquired deficiency<sup>20,53–55</sup>.

High LET radiation appeared to be a more potent carcinogen when compared to the same dose of gamma rays, causing early onset tumours and increased tumour burden in the majority of tissue types. While it should be cautioned that in the absence of a dose-response evaluation one cannot ascertain the relative biological equivalence of the doses of gamma versus Fe-ion radiation used in these experiments, these findings nevertheless agree with prior work showing increased colon tumourigenicity induced by Fe ions compared to gamma radiation in the APC/Min mouse model<sup>56</sup>. Heavy-ion exposure was also associated with an increased proportion of carcinomas, and despite equal numbers of animals receiving the two radiation qualities, the incidence of the two most commonly observed tumours (mammary and lung) was significantly different for  $^{56}\text{Fe}$ -ions compared to gamma radiation. Furthermore, *Trp53* genotype impacts tumour histology, as mice hemizygous for p53 were more likely to develop sarcomas and spindle tumours, independent of radiation quality. In contrast, the incidences of squamous cell carcinomas and lymphomas were 10-fold and 5-fold higher, respectively, in *Trp53 $\Delta$ P* versus *Trp53+/-* mice (Fig. 1c).

Our data on the patterns of genomic alterations in these different tumour types may help to rationalise these observations. On average, carcinomas exhibited higher numbers of complex chromosome aberrations, while sarcomas that arose in the same cohort of mice were genomically much more stable (Fig. 3c). Heavy ion exposure was strongly associated with an increased fraction of focal SVs, deletions and gene copy number changes relative to gamma exposure, particularly in mice carrying the *Trp53 $\Delta$ P* allele. As a result, the *Trp53 $\Delta$ P* genotype in combination with  $^{56}\text{Fe}$ -ion exposure resulted in the highest incidence of genomically unstable carcinomas. We conclude that heavy ion radiation exposure is, on average, more likely to generate the level of chromosomal instability required for transformation of the appropriate cell of origin for carcinomas.

Over 60% of mammary tumours, regardless of germline *Trp53* status, showed rearrangement and high-level amplification of the *Met* locus on chromosome 6. These tumours were predominantly induced by heavy ion radiation, which is compatible with the genome-wide observation of increased focal genomic events associated with exposure to  $^{56}\text{Fe}$ -ions. The propensity of heavy ion radiation to induce focal DNA damage, together with a specific feature of *Met* signalling in the mammary gland, may select for common tissue specific activation of this pathway.

Several examples were observed of tumours with complex patterns of whole chromosome or focal SVs, particularly in chromosome 1 and focal regions on chromosomes 6, 11, and 17. These features are similar to those generated by chromothripsis<sup>23</sup> in



**Fig. 6 Chromothripsis in radiation-induced tumours. a** Mapping of different somatic mutation types across chromosome 1 in all of the whole genome sequenced mammary tumours. Sub: Substitutions, ITX: Intrachromosomal translocation, INV: Inversion rearrangements, INS: Insertion rearrangements, DEL: Deletion rearrangements, CTX: Interchromosomal translocation, Ins: Small Insertions, Del: Small deletions, Complex: Complex INDELS. **b** Sample MD5207c has a strikingly large number of mutations of different types spread widely across the length of the chromosome, indicative of whole chromosome chromothripsis.

human tumours. A strong interaction between germline *Trp53* status and radiation quality in determining the genomic changes in tumours is exemplified by analysis of chromosome 11, which harbours the *Trp53* gene. The simplest way to lose *Trp53* function completely in *Trp53*<sup>+/-</sup> mice is to lose the wild type allele, by chromosome non-disjunction, or by more focal events which can be induced by heavy ion radiation (Fig. 7). The complete loss of function in *Trp53ΔP* mice is less likely to occur due to the presence of two partially functional alleles on different chromosomes. Simple non-disjunction would not lead to loss of function, and we only saw loss of *Trp53ΔP* alleles in tumours induced by <sup>56</sup>Fe-ions (Fig. 7), presumably due to the increased probability of focal DNA damage at the *Trp53* locus. These data emphasise the importance of interactions between inherited germline alleles and environmental exposures in determining tumour genome architecture.

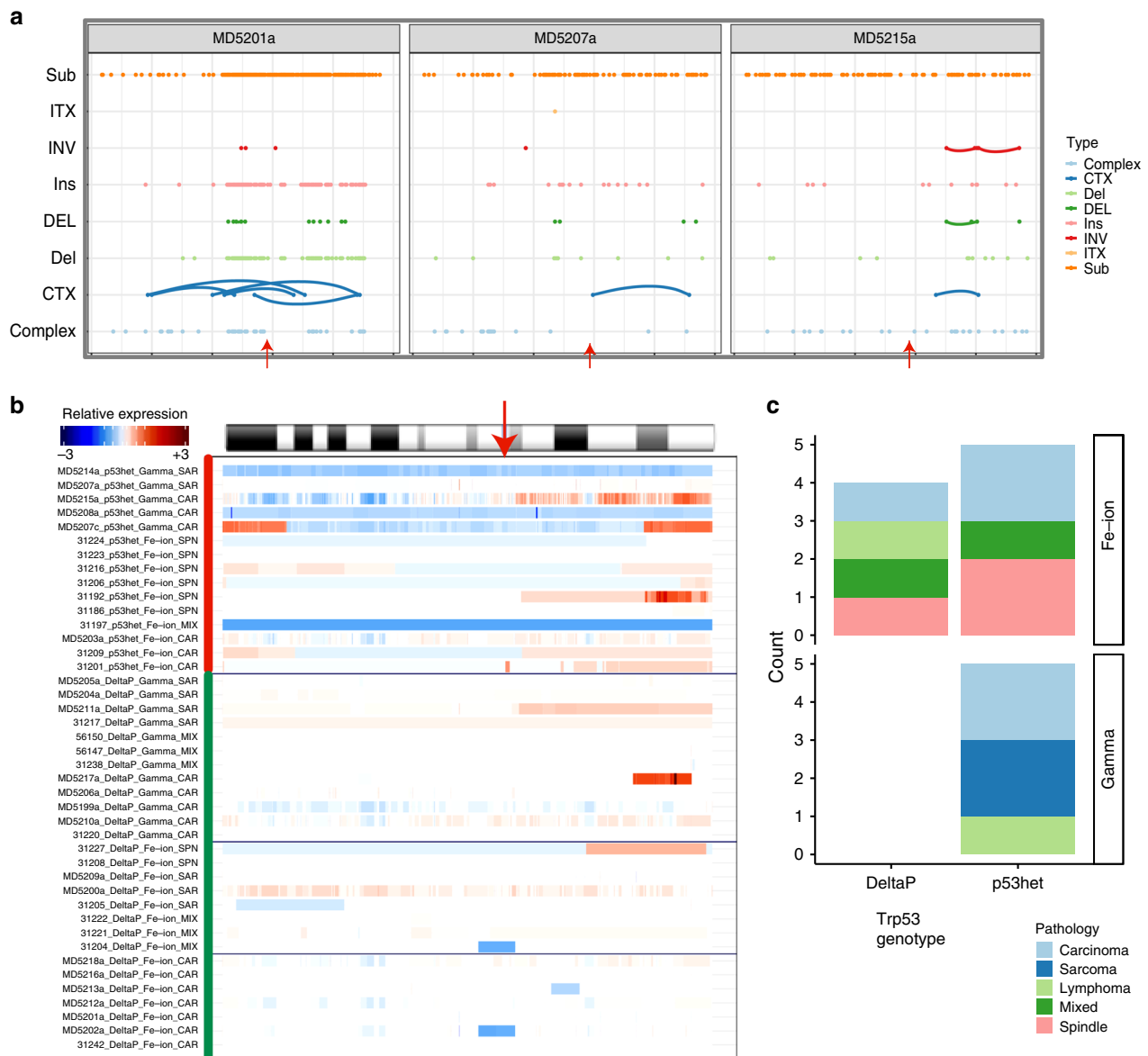
Finally, our data shed light on a long-standing question regarding mechanisms of DNA damage induced by ionizing radiation. Double strand breaks can be generated by a direct interaction between radiation particles and DNA, or indirectly by generation of hydroxyl radicals, leading to formation of oxidised bases and DNA single or double strand breaks. Deconvolution of the mutational signatures in mammary tumours identified a significant contribution from several known human signatures, including signature 18 attributed to the generation of ROS<sup>39,40</sup>.

Although the absolute number of point mutations attributable to this signature is relatively low, it was more prevalent in those tumours induced by gamma exposure compared to high energy <sup>56</sup>Fe-ions ( $P < 0.0318$ , Fig. 2b). High energy radiation, on the other hand, is more likely to damage DNA directly causing complex focal genomic changes leading to a higher level of genetic instability and more aggressive tumours. While there are some parallels between this work on mouse models and sequence analysis of radiation-associated human cancers<sup>4</sup>, detailed comparisons will require a much more complete analysis of the genetic alterations in human cancers that are more definitively linked to radiation of different qualities and exposure levels.

**Methods**

**Colony maintenance and breeding.** The *Trp53ΔP* allele was initially maintained on a mixed 129/SvJ and C57BL/6J background as has been described previously<sup>57</sup>. We bred the *Trp53ΔP* allele from the mixed 129/Sv and C57BL/6 background onto the FVB/N background with greater than 10 crosses resulting in 99.9% FVB/N background for these mice. The *Trp53*<sup>+/-</sup> allele has been previously bred onto the FVB background<sup>35</sup>. All mouse experiments received ethical approval by the University of California San Francisco Laboratory Animal Resource Centre. We have complied with all relevant ethical regulations for animal testing and research.

**Irradiation.** To assess the effect of dose and compare HI to gamma radiation, we exposed mixed (129/C57) and pure (FVB/N) background mice to radiation. Mice



**Fig. 7** Interaction between *Trp53* genotype and RT quality in complex structural variants. **a** Three samples showing representative enrichment of SVs on chromosome 11 near the *Trp53* locus, some with evidence of co-localisation of breakpoints and/or variants. Arrows denote the location of the *Trp53* gene along chromosome 11. **b** Tumours from *Trp53*<sup>+/-</sup> mice show common non-focal deletions or whole-chromosomal aneuploidy in chr 11 (top panel). These large scale copy number alterations are less common in tumours from the *Trp53* $\Delta$ P mice (lower panels). Arrow denotes the location of *Trp53*. **c** Effect of *Trp53* germline genotype on patterns of somatic alterations at the *Trp53* gene in mammary tumours. *Trp53* deletions are frequently observed in animals of *Trp53* heterozygous background exposed to either Fe ions or gamma radiation. In *Trp53* $\Delta$ P mice, localised deletions at the *Trp53* locus were only seen in tumours induced by <sup>56</sup>Fe ions.

were between 6 and 10 weeks of age. Mice exposed to HI radiation were subjected to 29 cGy, 50 cGy, 81 cGy or 1 Gy of 600 MeV/amu <sup>56</sup>Fe-ions at a dose rate of 50 cGy/min delivered by a 20 × 20 cm beam. Gamma radiation consisted of 50 cGy or 1 Gy of <sup>137</sup>Cs gamma rays at a dose rate of 25 cGy/min. Irradiation was performed at the Brookhaven national laboratory. Mice were shipped from UCSF to Brookhaven National Laboratories and rested for at least 3 day prior to whole-body irradiation. Following irradiation, mice were returned to UCSF for monitoring. Sham irradiated mice were flown to Brookhaven but were not irradiated.

**Tumour analysis and pathology.** Mice were monitored for the development of tumours at UCSF by both visual evaluation and palpation for a minimum of 500 days, until the mouse body-condition score deteriorated, or a tumour exceeded 1 cm in diameter.

At the time of sacrifice radiosensitive tissues including the skin, mammary, thymus and spleen were harvested along with the tumour and the tail as a non-radiation sensitive control. A small section of each tissue was preserved for

histology and the remainder was frozen for DNA, RNA or protein extraction. H&E slides were prepared from tumour samples and examined by a board-certified pathologist to determine histological subtypes. Tumours were categorised based on histology and disease site at time of necropsy. Mammary tumours include carcinomas, spindle cell tumours, sarcomas and angiosarcomas, the latter which was verified by immunohistochemistry (CD31+). To accurately represent the range of mammary tumours identified, whole exome sequencing analysis was performed on all mammary tumour types. Only carcinomas and sarcomas (angiosarcomas were excluded) were used for whole genome sequencing.

**DNA and RNA extraction.** DNA was extracted from snap frozen tumour tissues using the Qiagen DNeasy kit as per manufacturer's instructions. Concentration and quality were determined by Nanodrop spectrophotometry and by PicoGreen (Invitrogen). RNA was extracted from snap frozen tissues using the Zymo Quick-RNA column purification kit. RNA concentration and quality were determined by Nanodrop and Agilent RNA Nano Bioanalyzer kit.

**Whole exome sequencing.** Exonic DNA was captured using the Agilent whole exome capture kit (SureSelect Mouse All Exon). Captured material was indexed and sequenced on the Illumina platform at the Wellcome Trust Sanger Institute at a minimum depth of 100×. Raw 75 bp pair end sequencing reads were aligned with BWA-aln (v0.5.10)<sup>58</sup> to the GRCm38 mouse reference genome and were subsequently realigned around known InDels using GATK InDel re-aligner (v1.5-9) producing a single Binary Alignment Mapping (BAM) file for each sample.

**Whole genome sequencing.** DNA was sheared to 450 bp fragment size using a Covaris S2. A single Illumina TruSeq v3 sequencing library was created for each strain according to manufacturer's protocols. Each library was sequenced on a HiSeq 2000 over four lanes. Sequencing reads from each lane were aligned to the C57BL/6 J GRCm38 (mm10) mouse reference genome using BWA-MEM (0.7.13). For each library, aligned reads from each lane were merged using Picard Tools (v1.64) resulting in a single BAM file for each sample with a median coverage of 39×.

**SNV analysis.** SNV calling for exomes was performed using MuTect (version 1.1.6)<sup>59</sup>. Tumour samples were called against two genotype and strain matched controls and filtered for quality. The following filters were used: minimum of 10 q20 or greater covering bases in both tumour and normal, minimum tumour alternate allele fraction of 0.2, minimum of 3 alternate alleles in tumour, no more than 3 alternate alleles in either normal. Variants were also filtered for known variation using the Sanger variant dataset (mgp.v3.snps.rsIDdbSNPv137, filtered to exclude wild mice and all variants with *q* less than 20), and a pooled set of variants and variable regions observed in all control samples (called by MuTect2 version 3.6)<sup>59</sup>.

Evaluating the result revealed the existence of a large number of likely germline variants in non-coding regions. To remove these variants, we selected all variants that were exactly replicated (same location, same alternate allele) in at least three samples without a consequential coding effect. In all cases these variants appeared to be approximately heterozygous, and no variant was consequential. As a result, we omitted all variants exactly replicated in three or greater samples without an effect on protein sequence, and all variants exactly replicated seven or more times without regard to context.

Initial analysis revealed the existence of a large number of variants with the motif GGTGN. These sites were overwhelmingly associated with T > G transversions resulting in a GGGG homopolymer. Sanger sequencing showed that the majority of these variants were false positives. We imposed an additional, substantially more stringent, set of filters for these sites: minimum of 50 q20 or greater covering bases in both tumour and normal, minimum tumour alternate allele fraction of 0.2, minimum of 15 alternate alleles in tumour, normal alternate allele fraction of no greater than 0.05. This dramatically reduced the influence of this artifact. After this filtration, we took the intersection of the pass-filter variants between both controls as the keeper set for each tumour.

For WGS, somatic variants were detected using CaVEMan, an expectation maximisation-based somatic substitution detection algorithm<sup>60</sup>. Detected somatic variants were then filtered using an array of quality filters and common mouse genome variants were excluded. Tumours were called against paired tail DNA and filtered for the following: greater than or equal to 10 q20 or more bases, alternate allele fraction greater than 0.1 in tumour, and alternate allele fraction less than 0.1 in normal. In addition, we removed all variants occurring in simple repeat regions (simpleRepeat.txt, downloaded from [hgdownload.cse.ucsc.edu/goldenpath/mm10/database/](http://hgdownload.cse.ucsc.edu/goldenpath/mm10/database/))<sup>61</sup>.

**Indel analysis.** Indel calling in exome samples was performed using pindel (version 0.2.4w)<sup>62</sup>. Samples were called against two normal with the following options: -l false -x 1 -M 4. Results were then filtered for germline events, recurrence, and simple repeats. Germline filtration was accomplished by identifying all sites found in controls in this cohort and flagging any tumour indel that overlapped (within 2 bases) these sites. We also repeated this strategy for germline events found in the Sanger list of indel events (mgp.v3.indels.rsIDdbSNPv137, filtered for minimum quality 20 and no variants present only in wild derived strains).

We noticed a large degree of exact recurrence of indels within the tumour cohort, even following germline filtration. After manual examination, we determined the majority of these variants were likely to be germline or the result of recurrent sequencing/calling artifacts. To remove these variants, we removed all non-consequential indels that were exactly replicated (same location and alternate allele) in 3 or more samples. We also flagged regions (50 bp windows) containing three or more indels of any type. Manual evaluation of these regions revealed them to be primarily artefactual, with a small number of potentially biologically meaningful sites. As the number of artefactual variants substantially outnumbered the number of biologically meaningful sites, we omitted these regions for the calculation of indel rates but included them for the analysis of functional impact on cancer associated genes. We also removed all exome indels in simple repeat regions.

Indels were called in whole-genome samples by comparing tumour to paired control tail DNA using a modified version of the pindel programme<sup>62</sup>. Resulting lesions were then filtered for exact and regional recurrence (more than three exact replicates or indels within 50 bp across the tumour cohort), and presence within simple repeat regions. Indel rates were defined similar to SNVs.

**Mutation signature analysis.** The mutational catalogues from the 21 WGS samples were analyzed for mutational signatures. Signature extraction was performed using SigProfiler (conducted using methods based on nonnegative matrix factorisation (NMF) as described in Alexandrov et al. (refs. 1,36,37)). Analyses were carried out separately for single base substitutions (SBS signatures) and indels (ID signatures). We first performed de novo mutation signature extraction using NMF-based method as described in SigProfiler.

For SBS signatures, extracted signatures were then compared to the set of mutational signatures deciphered from the COSMIC database as previously described (refs. 1,36,37). The algorithm identifies the optimal combination of known human signatures that explains the observed mutation patterns (highest cosine similarity). All signatures extracted across all samples were able to be explained by a combination of signatures from the known human signatures (cosine similarity > 0.75) and thus we did not identify any signatures that would be considered novel. We performed hierarchical clustering of the samples based on the relative contribution of identified mutation signatures using both results from the de novo extraction and after signature deconvolution.

For INDELS, given the limited understanding for these signatures, particularly in mouse, as well as the fact that there were very few mutations, we chose to perform downstream analysis using only the de novo extracted signatures.

Code for the original SigProfiler software is available: <https://www.mathworks.com/matlabcentral/fileexchange/38724-sigprofiler>.

For further details of this algorithm and it's most up to date implementation we refer the reader to ref. 1 and the updated software webpage: <https://pypi.org/project/sigprofiler/>.

**CNV segmentation and analysis.** CNV segmentation in exomes was performed using CNVkit<sup>63</sup> following the default workflow with target regions defined from the UCSC refgene flatfile gene locations padded by +/-500 basepairs.

To remove spuriously aberrant regions we assessed control copy-number profiles for aneuploidies. Any region found to be aberrant in two or more controls was omitted from the tumour samples.

CNV segmentation in whole-genome samples was performed using BICseq (version 1.1.2)<sup>64</sup> with the following options: --paired --l = 400,50 --bin\_size = 250. Three samples showed consistently patterned aneuploidies that appeared to be artefactual and were omitted (MD5215a, MD5212a, MD5210a).

Amplification was defined as a copy-number log<sub>2</sub> ratio over 0.15, and deletion as log<sub>2</sub> ratio less than -0.15. A whole-chromosome amplification was defined as greater than 50% of the chromosome covered by copy-number bins, and greater than 40% of the chromosome called as amplified. Whole-chromosome deletions were defined with similar proportions. Mixed chromosomes were chromosomes that fit the definition of amplification or deletion, but with 5% or more of the chromosomes total length aneuploid in the opposite direction. This definition captured the majority of the visibly mixed chromosomes without including any discernable false positives.

Significantly recurrent whole-chromosome aneuploidies were identified by first permuting chromosome state per sample 1000 times, counting the maximum observed quantity of aneuploidies per chromosome in the randomised data, and identifying the threshold that was greater than 95% of observed maximum aneuploidy counts in the permuted data. This was done for any aneuploidy, amplifications only, and deletions only.

Focal amplifications and deletions were identified through the cghMCR package<sup>65</sup> with amplification and deletion thresholds increased to +/-0.25 to focus on high-level aneuploidies. Significance threshold was defined by 1000 permutations as above.

**Rearrangement analysis.** Rearrangements were called using both lumpy<sup>66</sup> and breakdancer<sup>67</sup> but only breakdancer calls were used due to extremely high agreement between the intrachromosomal events from both callers, and the high level of spurious interchromosomal rearrangements identified by lumpy. These spurious calls appeared to corroborate recent findings showing that lumpy performed poorly when calling non-C57B6 samples against the C57BN6 genome<sup>68</sup>. Breakdancer was used with the following arguments: -r 10 -q 30.

Rearrangements were filtered by depth and region. Depth filters included the following: less than 2 control reads, and greater than or equal to 10 tumour reads. Regional filters included simple repeats and mapability. Mapability was defined as uniquely mapping from the crgMapabilityAlign50mer.txt file for the mm9 genome (downloaded from [hgdownload.cse.ucsc.edu/goldenpath/mm9/database/](http://hgdownload.cse.ucsc.edu/goldenpath/mm9/database/)) remapped to mm10 with liftOver<sup>61</sup>. Any rearrangement with either endpoint within a simple repeat or non-uniquely mapping region was omitted.

Focal rearrangement hotspots were identified by establishing the intermutation distance (the distance between any given rearrangement and its nearest neighbour) in 1 Mb windows sliding by 100 kb intervals. Enriched sites were defined as those windows with intermutation distance greater than five standard deviations greater than the genome-wide mean, with at least six unique lesions, in at least three different samples.

**Statistical analysis and plotting.** Statistical analysis was performed in R (version 3.3.3)<sup>69</sup>. Two-group analyses were performed using the non-parametric Mann-Whitney U test. Fisher's test was used to assess categorical data. The *glm* function was used to model the relationship between whether or not a chromosome was

aneuploid and genotype/radiation exposure. Survival analysis was performed using Cox proportional hazards modelling and significance evaluated by the log-rank test. Plots were generated using the *ggplot2* package<sup>70</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The whole-exome sequencing data have been deposited in the ENA database under the accession code ERP001454 [<https://www.ebi.ac.uk/ena/data/view/PRJEB3044>]. The whole-genome sequencing data have been deposited in the ENA database under the accession code ERP001454 [<https://www.ebi.ac.uk/ena/data/view/PRJEB13771>]. All the other data supporting the findings of this study are available within the article and its supplementary information files and from the corresponding author upon request.

### Code availability

All software utilised for this paper are publicly available and are as indicated in the Methods section. Any custom code or scripts used in this paper are available upon request.

Received: 31 May 2019; Accepted: 17 December 2019;

Published online: 20 January 2020

### References

- Alexandrov, L. et al. The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* 322859, <https://doi.org/10.1101/322859> (2018).
- Westcott, P. M. K. et al. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2014).
- McCreery, M. Q. et al. Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. *Nat. Med.* **21**, 1514–1520 (2015).
- Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
- Zeitlin, C. & La Tessa, C. The role of nuclear fragmentation in particle therapy and space radiation protection. *Front. Oncol.* **6**, 65 (2016).
- Durante, M. New challenges in high-energy particle radiobiology. *Br. J. Radiol.* **87**, 20130626 (2014).
- Nguyen, D. H. et al. Radiation acts on the microenvironment to affect breast carcinogenesis by distinct mechanisms that decrease cancer latency and affect tumor type. *Cancer Cell* **19**, 640–651 (2011).
- Brenner, D. J. & Elliston, C. D. The potential impact of bystander effects on radiation risks in a mars mission. *Radiat. Res.* **156**, 612–617 (2001).
- Nagasawa, H. & Little, J. B. Induction of sister chromatid exchanges by extremely low doses of alpha-particles. *Cancer Res.* **52**, 6394–6396 (1992).
- Barcellos-Hoff, M. H. et al. Concepts and challenges in cancer risk prediction for the space radiation environment. *Life Sci. Sp. Res.* **6**, 92–103 (2015).
- Locke, P. A. & Weil, M. M. Personalized cancer risk assessments for space radiation exposures. *Front. Oncol.* **6**, 38 (2016).
- Durante, M. & Cucinotta, F. A. Heavy ion carcinogenesis and human space exploration. *Nat. Rev. Cancer* **8**, 465–472 (2008).
- Ohnishi, T. et al. Detection of space radiation-induced double strand breaks as a track in cell nucleus. *Biochem. Biophys. Res. Commun.* **390**, 485–488 (2009).
- Groesser, T., Chun, E. & Rydberg, B. Relative biological effectiveness of high-energy iron ions for micronucleus formation at low doses. *Radiat. Res.* **168**, 675–682 (2007).
- Durante, M. et al. Complex chromosomal rearrangements induced in vivo by heavy ions. *Cytogenet. Genome Res.* **104**, 240–244 (2004).
- Riley, P. A. Free radicals in biology: oxidative stress and the effects of ionizing radiation. *Int. J. Radiat. Biol.* **65**, 27–33 (1994).
- Lane, D. P. p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
- Muller, P. A. J. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304–317 (2014).
- Goldstein, M. & Kastan, M. B. The DNA damage response: implications for tumor responses to radiation and chemotherapy. *Annu. Rev. Med.* **66**, 129–143 (2015).
- Li, F. P. & Fraumeni, J. F. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann. Intern. Med.* **71**, 747–752 (1969).
- Li, F. P. & Fraumeni, J. F. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J. Natl Cancer Inst.* **43**, 1365–1373 (1969).
- Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
- Mardin, B. R. et al. A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
- Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Cortés-Ciriano, I. et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing on behalf of the PCAWG Structural Variation Working Group and the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Network. <https://doi.org/10.1101/333617> (2018).
- Adams, C. J. et al. The Trp53 delta proline (Trp53ΔP) mouse exhibits increased genome instability and susceptibility to radiation-induced, but not spontaneous, tumor development. *Mol. Carcinog.* **55**, 1387–1396 (2016).
- Kemp, C. J., Wheldon, T. & Balmain, A. p53-deficient mice are extremely susceptible to radiation-induced tumorigenesis. *Nat. Genet.* **8**, 66–69 (1994).
- Toledo, F. et al. A mouse p53 mutant lacking the proline-rich domain rescues Mdm4 deficiency and provides insight into the Mdm2-Mdm4-p53 regulatory network. *Cancer Cell* **9**, 273–285 (2006).
- Duan, X. & Li, J. Association between MDM2 SNP309, p53 Arg72Pro, and hepatocellular carcinoma risk. *Medicine* **96**, e7856 (2017).
- Xue, L. et al. MDM2 and P53 polymorphisms contribute together to the risk and survival of prostate cancer. *Oncotarget* **7**, 31825–31831 (2016).
- Kiffer, F., Boerma, M. & Allen, A. Behavioral effects of space radiation: a comprehensive review of animal studies. *Life Sci. Sp. Res.* **21**, 1–21 (2019).
- Wilson, J. W. et al. Exposures to solar particle events in deep space missions. *Tech. Report, NASA-TP-3668; NAS 1.603668; L-17616* (1997).
- Kim, M.-H. Y., Hayat, M. J., Feiveson, A. H. & Cucinotta, F. A. Prediction of frequency and exposure level of solar particle events. *Health Phys.* **97**, 68–81 (2009).
- Kirova, Y. M., Vilcoq, J. R., Asselain, B., Sastre-Garau, X. & Fourquet, A. Radiation-induced sarcomas after radiotherapy for breast carcinoma. *Cancer* **104**, 856–863 (2005).
- Mao, J.-H. et al. Genomic instability in radiation-induced mouse lymphoma from p53 heterozygous mice. *Oncogene* **24**, 7924–7934 (2005).
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
- Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Tominaga, H., Kodama, S., Matsuda, N., Suzuki, K. & Watanabe, M. Involvement of reactive oxygen species (ROS) in the induction of genetic instability by radiation. *J. Radiat. Res.* **45**, 181–188 (2004).
- Viel, A. et al. A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. *EBioMedicine* **20**, 39–49 (2017).
- Pilati, C. et al. Mutational signature analysis identifies *MUTYH* deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **242**, 10–15 (2017).
- Sharma, V. et al. Oxidative stress at low levels can induce clustered DNA lesions leading to NHEJ mediated mutations. *Oncotarget* **7**, 25377–25390 (2016).
- Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* **49**, 341–348 (2017).
- Fahrlander, P. D. et al. Activation of the c-myc oncogene by the immunoglobulin heavy-chain gene enhancer after multiple switch region-mediated chromosome rearrangements in a murine plasmacytoma. *Proc. Natl Acad. Sci. USA* **82**, 3746–3750 (1985).
- Zhang, C.-Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
- Kloosterman, W. P. & Cuppen, E. Chromothripsis in congenital disorders and cancer: similarities and differences. *Curr. Opin. Cell Biol.* **25**, 341–348 (2013).
- Mills, R. E. et al. Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
- Ratnaparkhe, M. et al. Genomic profiling of Acute lymphoblastic leukemia in ataxia telangiectasia patients reveals tight link between ATM mutations and chromothripsis. *Leukemia* **31**, 2048–2056 (2017).
- Egeblad, M. & Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**, 161–174 (2002).
- Jen, K.-Y. et al. Sequential mutations in Notch1, Fbxw7, and Tp53 in radiation-induced mouse thymic lymphomas. *Blood* **119**, 805–809 (2012).
- Sherborne, A. L. et al. Mutational analysis of ionizing radiation induced neoplasms. *Cell Rep.* **12**, 1915–1926 (2015).
- Davidson, P. R., Sherborne, A. L., Taylor, B., Nakamura, A. O. & Nakamura, J. L. A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies. *Sci. Rep.* **7**, 7645 (2017).
- Chang, H.-H. et al. Incidence of pancreatic cancer is dramatically increased by a high fat, high calorie diet in KrasG12D mice. *PLoS One* **12**, e0184455 (2017).
- Hodgson, D. C. et al. Long-term solid cancer risk among 5-year survivors of Hodgkin's lymphoma. *J. Clin. Oncol.* **25**, 1489–1497 (2007).

54. Behrens, C. et al. Molecular changes in second primary lung and breast cancers after therapy for Hodgkin's disease. *Cancer Epidemiol. Biomark. Prev.* **9**, 1027–1035 (2000).
55. Malkin, D. et al. Germline mutations of the p53 tumor-suppressor gene in children and young adults with second malignant neoplasms. *N. Engl. J. Med.* **326**, 1309–1315 (1992).
56. Datta, K., Suman, S., Kallakury, B. V. S. & Fornace, A. J. Heavy ion radiation exposure triggered higher intestinal tumor frequency and greater  $\beta$ -catenin activation than  $\gamma$  radiation in APCMin/+ Mice. *PLoS One* **8**, e59295 (2013).
57. Toledo, F. et al. Mouse mutants reveal that putative protein interaction sites in the p53 proline-rich domain are dispensable for tumor suppression. *Mol. Cell. Biol.* **27**, 1425–1432 (2007).
58. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
59. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
60. Jones, D. et al. cgpCaVEManWrapper: Simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. in *Current Protocols in Bioinformatics* **56**, 15.10.1–15.10.18 (2016).
61. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
62. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. 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 (2009).
63. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
64. Xi, R., Luquette, J., Hadjipanayis, A., Kim, T.-M. & Park, P. J. BIC-seq: a fast algorithm for detection of copy number alterations based on high-throughput sequencing data. *Genome Biol.* **11**, O10 (2010).
65. Zhang, J. & Feng, B. cghMCR: Find chromosome regions showing common gains/losses. <https://rdr.io/bioc/cghMCR/> (2018).
66. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
67. Fan, X., Abbott, T. E., Larson, D. & Chen, K. BreakDancer: identification of genomic structural variation from paired-end read mapping. *Curr. Protoc. Bioinforma.* **45**, 15.6.1–11 (2014).
68. Ahdesmäki, M. J. et al. Prioritisation of structural variant calls in cancer genomes. *PeerJ* **5**, e3166 (2017).
69. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2017).
70. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).

## Acknowledgements

This work was supported by a Cancer Research UK Grand Challenge Award [C98/A24032], US National Cancer Institute (NCI) grants RO1CA184510, UO1CA176287, R35CA210018 and the Barbara Bass Bakar Professorship of Cancer Genetics (to A.B.). The generation of radiation-induced mouse tumours was supported by a sub-award from

NASA grant NNXD 9AM52G (PI MH Barcellos-Hoff) and DE-SC0003679 from the DOE. The DNA sequencing was funded by a CRUK Program Grant to D.J.A. Y.R.L. was supported by NCI F32 CA 232635. K.D.H. was supported by NCI U01 CA 84244, NCI F31 CA 180715 and NIH T32 GM 7175-35. We thank Professor Mike Stratton for comments and support.

## Author contributions

The project was conceived and supervised by A.B. The paper was written by Y.-R.L., K.H., C.A. and A.B., with contributions from other authors. Y.-R.L. and K.H. carried out computational analysis of whole genome and exome sequencing. Data analysis and interpretation was carried out by Y.-R.L., K.H., C.A., L.A. and A.B. C.A. carried out all animal studies with help from R.D. DNA sequencing was carried out at the Sanger Institute under the supervision of D.J.A. V.I., L.R. and R.M. assisted with the sequencing pipeline, and K.-Y.J. carried out histopathological analysis of tumour samples. E.F. provided assistance with computational and statistical analyses, and G.H. coordinated sample selection, database and tissue bank tracking.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-14261-4>.

Correspondence and requests for materials should be addressed to D.A. or A.B.

Peer review information *Nature Communications* thanks Simon Powell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020