



# Mutational signatures of ionizing radiation in second malignancies

## Citation

Behjati, S., G. Gundem, D. C. Wedge, N. D. Roberts, P. S. Tarpey, S. L. Cooke, P. Van Loo, et al. 2016. "Mutational signatures of ionizing radiation in second malignancies." *Nature Communications* 7 (1): 12605. doi:10.1038/ncomms12605. <http://dx.doi.org/10.1038/ncomms12605>.

## Published Version

doi:10.1038/ncomms12605

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:29407602>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

ARTICLE

Received 7 Sep 2015 | Accepted 13 Jul 2016 | Published 12 Sep 2016

DOI: 10.1038/ncomms12605

OPEN

# Mutational signatures of ionizing radiation in second malignancies

Sam Behjati<sup>1,2,\*</sup>, Gunes Gundem<sup>1,\*</sup>, David C. Wedge<sup>1,3</sup>, Nicola D. Roberts<sup>1</sup>, Patrick S. Tarpey<sup>1</sup>, Susanna L. Cooke<sup>1</sup>, Peter Van Loo<sup>4,5</sup>, Ludmil B. Alexandrov<sup>1</sup>, Manasa Ramakrishna<sup>1</sup>, Helen Davies<sup>1</sup>, Serena Nik-Zainal<sup>1</sup>, Claire Hardy<sup>1</sup>, Calli Latimer<sup>1</sup>, Keiran M. Raine<sup>1</sup>, Lucy Stebbings<sup>1</sup>, Andy Menzies<sup>1</sup>, David Jones<sup>1</sup>, Rebecca Shepherd<sup>1</sup>, Adam P. Butler<sup>1</sup>, Jon W. Teague<sup>1</sup>, Mette Jorgensen<sup>6</sup>, Bhavisha Khatri<sup>7</sup>, Nischalan Pillay<sup>6,7</sup>, Adam Shlien<sup>1,8</sup>, P. Andrew Futreal<sup>1,9</sup>, Christophe Badie<sup>10</sup>, ICGC Prostate Group<sup>‡</sup>, Ultan McDermott<sup>1</sup>, G. Steven Bova<sup>11</sup>, Andrea L. Richardson<sup>12,13,†</sup>, Adrienne M. Flanagan<sup>6,7</sup>, Michael R. Stratton<sup>1</sup> & Peter J. Campbell<sup>1,14</sup>

Ionizing radiation is a potent carcinogen, inducing cancer through DNA damage. The signatures of mutations arising in human tissues following *in vivo* exposure to ionizing radiation have not been documented. Here, we searched for signatures of ionizing radiation in 12 radiation-associated second malignancies of different tumour types. Two signatures of somatic mutation characterize ionizing radiation exposure irrespective of tumour type. Compared with 319 radiation-naïve tumours, radiation-associated tumours carry a median extra 201 deletions genome-wide, sized 1–100 base pairs often with microhomology at the junction. Unlike deletions of radiation-naïve tumours, these show no variation in density across the genome or correlation with sequence context, replication timing or chromatin structure. Furthermore, we observe a significant increase in balanced inversions in radiation-associated tumours. Both small deletions and inversions generate driver mutations. Thus, ionizing radiation generates distinctive mutational signatures that explain its carcinogenic potential.

<sup>1</sup>Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA UK. <sup>2</sup>Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. <sup>3</sup>Oxford Big Data Institute and Oxford Centre for Cancer Gene Research, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. <sup>4</sup>The Francis Crick Institute, London WC2A 3LY, UK. <sup>5</sup>Department of Human Genetics, University of Leuven, Leuven B-3000, Belgium. <sup>6</sup>University College London Cancer Institute, Huntley Street, London WC1E 6BT, UK. <sup>7</sup>Histopathology, Royal National Orthopaedic Hospital NHS Trust, Stanmore, Middlesex HA7 4LP, UK. <sup>8</sup>Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8. <sup>9</sup>Department of Genomic Medicine, MD Anderson Cancer Center, University of Texas, Houston, Texas 77030, USA. <sup>10</sup>Cancer Mechanisms and Biomarkers Group, Radiation Effects Department, Centre for Radiation Chemical and Environmental Hazards, Public Health England, Chilton, Didcot OX11 0RQ, UK. <sup>11</sup>Institute of Biosciences and Medical Technology, BioMediTech, University of Tampere and Fimlab Laboratories, Tampere University Hospital, Tampere FI-33520, Finland. <sup>12</sup>Dana-Farber Cancer Institute, Boston, Massachusetts 02215-5450, USA. <sup>13</sup>Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115 USA. <sup>14</sup>Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 2XY, UK. \* These authors contributed equally to this work. † Present address: Sibley Memorial Hospital, Johns Hopkins Medicine, Washington, District Of Columbia 20016, USA. Correspondence and requests for materials should be addressed to P.J.C. (email: pc8@sanger.ac.uk).

‡ A full list of consortium members appears at the end of the paper.

Exposure to ionizing radiation increases the risk of subsequent cancer. This risk exhibits a strong dose–response relationship, and there appear to be no safe limits for radiation exposure<sup>1</sup>. This association was first noted by March who observed an increased incidence of leukaemia amongst radiologists<sup>2</sup>. A leading cause of radiation-induced cancers appears to be exposure to medical radiation, either in the form of radiotherapy for an unrelated malignancy<sup>3</sup> or diagnostic radiography<sup>4,5</sup>. These iatrogenic tumours arise as *de novo* neoplasms in a field of therapeutic radiation after a latency period that can span decades<sup>6</sup>, and are not recurrences of the original cancer<sup>7</sup>.

Many, but not all, environmental carcinogens induce cancer by increasing the rate of mutation in somatic cells. The physico-chemical properties of a given carcinogen govern its interaction with DNA, leading to recurrent ‘signatures’ or patterns of mutations in the genome. These can be reconstructed either from experimental model systems<sup>8,9</sup> or from statistical analyses of cancer genomes in exposed patients<sup>10–12</sup>. Ionizing radiation directly damages DNA, and can generate lesions on single bases, single-stranded nicks in the DNA backbone, clustered lesions at several nearby sites and double-stranded DNA breaks<sup>13</sup>. In experimental systems exposed to radiation, including the murine germline and *Arabidopsis thaliana* cells, ionizing radiation can cause all classes of mutations, with possible enrichment of indels<sup>14–22</sup>. Targeted gene screens in radiation-induced sarcoma have indicated an increased burden of deletions and substitutions with frequent inactivation of *TP53* and *RB1* (refs 23–25). In addition, a transcriptome profile that represents a state of chronic oxidative stress has been proposed to be specific to radiation-associated sarcoma<sup>26</sup>.

We studied the genomes of 12 radiation-associated second malignancies of four different tumour types: osteosarcoma; spindle cell sarcoma; angiosarcoma; breast cancer. These were secondary tumours that arose within a field of therapeutic ionizing radiation and were not thought to be recurrences of the original malignancy treated with radiation. We chose this experimental design for several reasons: the tumours are classic radiotherapy-induced cancers with high attributable risks for the radiation exposure; the radiation exposure occurs over a short time period relative to the evolution of the cancer; and the mutational signatures of sporadic breast cancers and sarcomas have been well documented<sup>10,27–29</sup>. It should be noted that in the absence of biomarkers, a diagnosis of a tumour being radiation-induced cannot be definitively made (see Supplementary Note 1 for clinical details and further discussion).

We subjected these 12 tumours, along with normal tissues from the same patients, to whole-genome sequencing and obtained catalogues of somatic mutations. We compared our findings to 319 radiation-naïve breast cancers and sarcomas processed by the same sequencing and bioinformatics pipeline: 251 breast tumours; 33 breast tumours with pathogenic *BRCA1* or *BRCA2* germline mutations; 35 osteosarcomas (see Methods for cohort details). In addition, we validated our findings in a published series of radiation-naïve and radiation-exposed prostate tumours from ten patients<sup>30</sup>.

The main aim of our analyses was to search for tumour-type independent, overarching signatures of ionizing radiation. Overall we identified two such signatures in radiation-associated second malignancies, an excess of balanced inversions and of small deletions.

## Results

**Tumour-type specific features.** The 12 radiation-associated tumours harboured 1,506–9,245 substitutions per genome

(median 4113), 135–943 indels per genome, (median 429) and 6–321 rearrangement break points per genome (median 74; Supplementary Data 1–3). The observed driver mutations followed the patterns expected for the tumour type consistently (Supplementary Table 1). Angiosarcomas harboured *PTPRB* and *PLCG1* mutations. In spindle cell sarcoma and osteosarcoma driver alteration of *TP53* and *CDKN2A* were found. Canonical *PIK3CA* mutations were seen in the radiation-associated breast tumours. Similarly, many of the mutational signatures seen in sporadic cancers were also present in radiation-associated tumours, such as chromothripsis in sarcomas (Supplementary Fig. 1). Against this backdrop of genomic diversity, we found evidence for two mutational signatures in the radiation-associated cancers that transcended tumour type: small deletions and balanced inversions.

### Enrichment of deletions in radiation-associated tumours.

Although the absolute burden of indels varied across the 12 radiation-associated tumours, in each tumour the indel burden was high compared with that tumour’s substitution burden (Fig. 1a). Compared with 319 radiation-naïve tumours, the indel/substitution ratio was significantly increased in radiation-associated tumours ( $P = 0.0003$ , linear mixed effects model, see Methods).

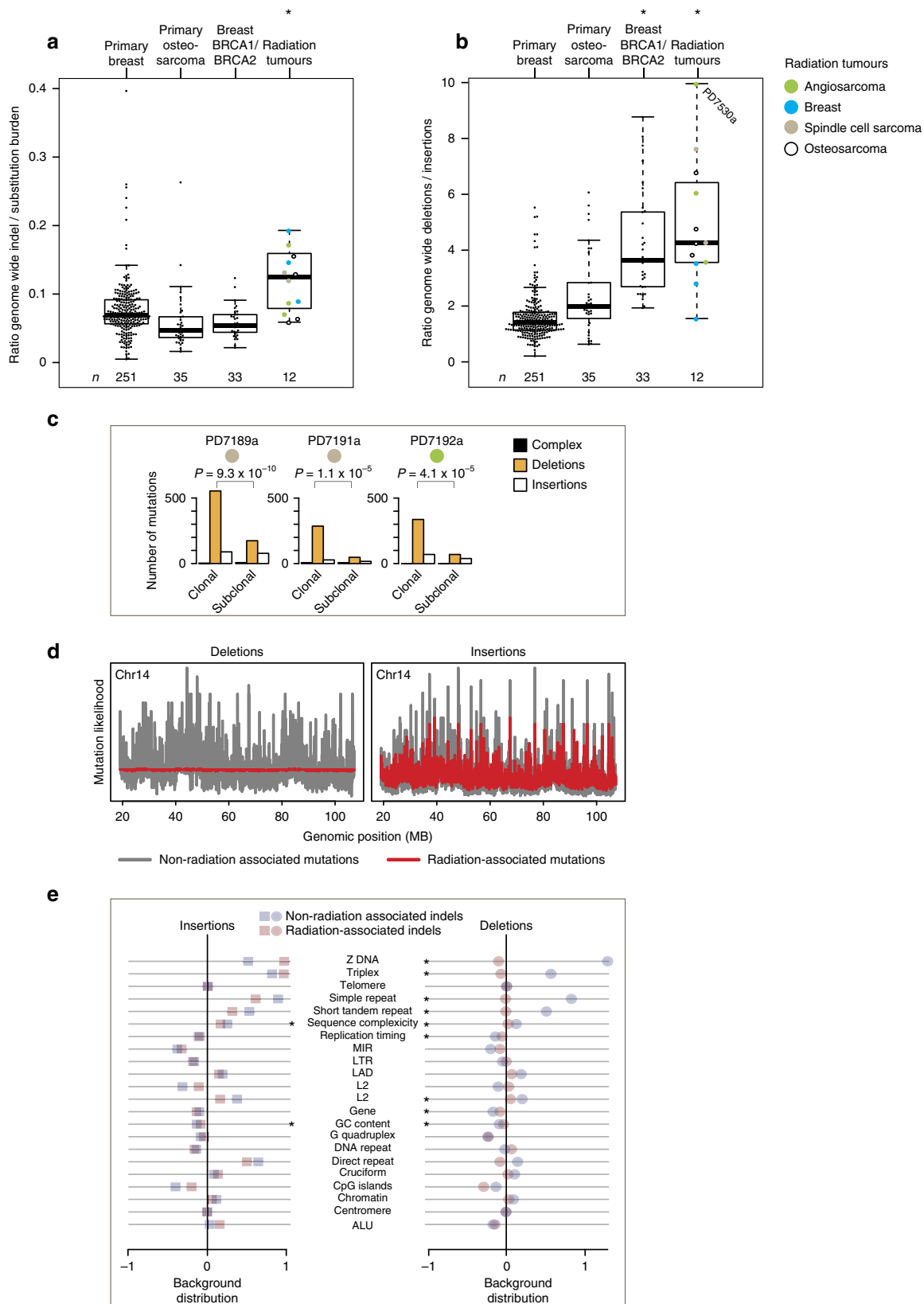
Deletions and insertions were not equally enriched. There was a significant excess of deletions relative to insertions in radiation-associated second malignancies (Fig. 1b;  $P < 2.2 \times 10^{-16}$ , linear mixed effects model). This excess of deletions was also seen in *BRCA1* or *BRCA2* germline-deficient breast tumours, as previously described<sup>27</sup>, but was not seen in radiation-naïve sporadic breast tumours or sarcomas. In each radiation-associated tumour, the radiotherapy had been given over a relatively short time period many years earlier. If the excess deletions we observed were directly attributable to ionizing radiation, then the enrichment should only be evident amongst the early, clonal mutations and not in the late, subclonal mutations. We were able to define subclones in three of the radiation-associated tumour genomes. Strikingly, compared with subclonal mutations, deletions were significantly increased compared with insertions amongst clonal mutations in all three cases (Fig. 1c,  $P < 0.00005$ ; Fisher’s exact test).

**Distribution of deletions across the genome.** Many mutational signatures show uneven distribution across the genome, especially those associated with carcinogens such as tobacco smoke and ultraviolet light<sup>31</sup>, thought to arise due to higher order chromatin organization and accessibility of the carcinogen and repair proteins to DNA (ref. 32). Deletions found in radiation-naïve tumours showed considerable long-range variation in density across the genome, as did insertions in both radiation-associated and radiation-naïve tumours, correlating with several genomic features (Fig. 1d,e; Supplementary Data 4). In stark contrast, deletions in radiation-induced cancers showed almost no variability across the genome and minimal correlation with genomic properties such as replication timing, sequence complexity or GC content. We hypothesize that this is because of the pervasive penetration of ionizing radiation through tissue, meaning that its interaction with DNA is stochastic and unaffected by higher order chromatin structure. Since small deletions are the predominant read-out of this damage, they show no association with the genomic features that influence other mutational processes.

**Evidence of non-homologous end-joining causing deletions.** In two aspects, the deletions of radiation-associated cancers

resembled those of *BRCA1* or *BRCA2* germline-deficient breast tumours (Supplementary Fig. 2): enrichment for deletions > 2–3 bp in length and significantly higher rates of microhomology at the breakpoint junction ( $P = 2 \times 10^{-16}$ , Kolmogorov–Smirnov test)<sup>27</sup>. This similarity suggests that microhomology mediated or

non-homologous end-joining are the pathways for repairing radiation-induced DNA damage, rather than homologous recombination. Possible explanations for this include that damage occurs at phases of the cell cycle when homologous recombination pathways are less active<sup>33</sup> or because the damaged



DNA ends contain structures that interfere with homologous recombination<sup>34</sup>.

**Enrichment of balanced inversions.** For structural variants, we found enrichment of a rare type of rearrangement, balanced inversions, among radiation-associated second malignancies, irrespective of tumour type (Table 1; Fig. 2). While rearrangements with an inverted orientation are common in cancer genomes, they are typically unbalanced, associated with copy-number changes and caused by processes such as breakage-fusion-bridge cycles<sup>35</sup>, chromothripsis<sup>29</sup> and chromoplexy<sup>36</sup>. We found a significant enrichment of balanced inversions in radiation-associated cancers: 52 in 11/12 tumours compared with 66 balanced inversions in 43 of the 286 radiation-naïve tumours studied ( $P=2 \times 10^{-16}$ , generalized linear model). Of note, complete inversions were also significantly enriched amongst *BRCA1* and *BRCA2* germline-deficient breast tumours (Table 1,  $P=2 \times 10^{-16}$ ). Balanced inversions ranged in size from a few hundred base pairs to nearly 100 megabases, and at breakpoint junctions showed variability in microhomology and in non-templated sequence inserted (Supplementary Data 5).

**Validation of findings in prostate tumours.** To validate our observation that deletions and balanced inversions are genomic imprints of ionizing radiation, we examined the genomes of primary and/or metastatic prostate tumours from ten patients, previously published<sup>30</sup>. Five patients had developed metastases after irradiation of the primary lesion; four patients had never received radiation treatment; and one patient, PD11331, received radiotherapy to the primary lesion after metastases had already formed. Consistent with the observations made in the 12 radiation-associated second malignancies, we found a significant enrichment of deletions in prostate cancer lesions exposed to radiotherapy compared with radiation-naïve tumours' ( $P=0.0002$ , generalized linear model, Fig. 3a). Strikingly, in patient PD11331, the radiation-exposed primary tumour (sample PD11331c), but not radiation-naïve metastases, exhibited a preponderance of deletions (Fig. 3b,  $P=10^{-15}$ , Fisher's exact test). Similarly, balanced inversions were enriched amongst radiation-exposed lesions ( $P=0.04$ , generalized linear model; Supplementary Table 2). In patient PD11331, again it was the radiation-exposed primary tumour, but not any of the metastases, that harboured a balanced inversion.

**Driver events generated by deletions and inversions.** The oncogenic potential of a mutational process derives from its

capacity to generate driver mutations. With their absence of copy-number effects, functional consequences of balanced inversions most commonly arise from genes broken at either end of the inversion, notwithstanding the possibility of long-range gene-enhancer disruption. In our data, 48/104 inversion break points disrupted or fused genes (Supplementary Data 5), with one forming a driver mutation through disruption of *TP53*. For the mutational signature of small deletions that we observe, we estimate that the median excess of indels in the radiation-induced cancers sequenced here is 201 indels per genome (linear mixed effects model, s.d. 348 indels). Among these are a 14 base pair deletion in *CASP8* and a 4 base pair deletion in *TP53*, both disrupting essential splice sites and thus generating driver events.

## Discussion

Overall we identified two genomic imprints of ionizing radiation, an excess of deletions and of an exceedingly rare type of rearrangement, balanced inversions. The validity of our study may be limited by the overall number of tumours we examined and the small number of each tumour type. Yet it would seem unlikely that the enrichment in radiation-associated tumours of deletions and of balanced inversions occurred by chance. This view is supported by our statistical analyses as well as the fact that the signatures were tumour-type independent. Both signatures were present across four different tumour types and could be validated in a cohort of radiation-exposed prostate cancer lesions, despite differences in the biological context of radiation-exposed prostate tumours and radiation-associated second malignancies (Supplementary Note 2). Particularly striking is patient PD11331 whose primary prostate lesion was irradiated after metastases had formed. The primary lesion, but not the metastases, exhibited the genomic features of ionizing radiation.

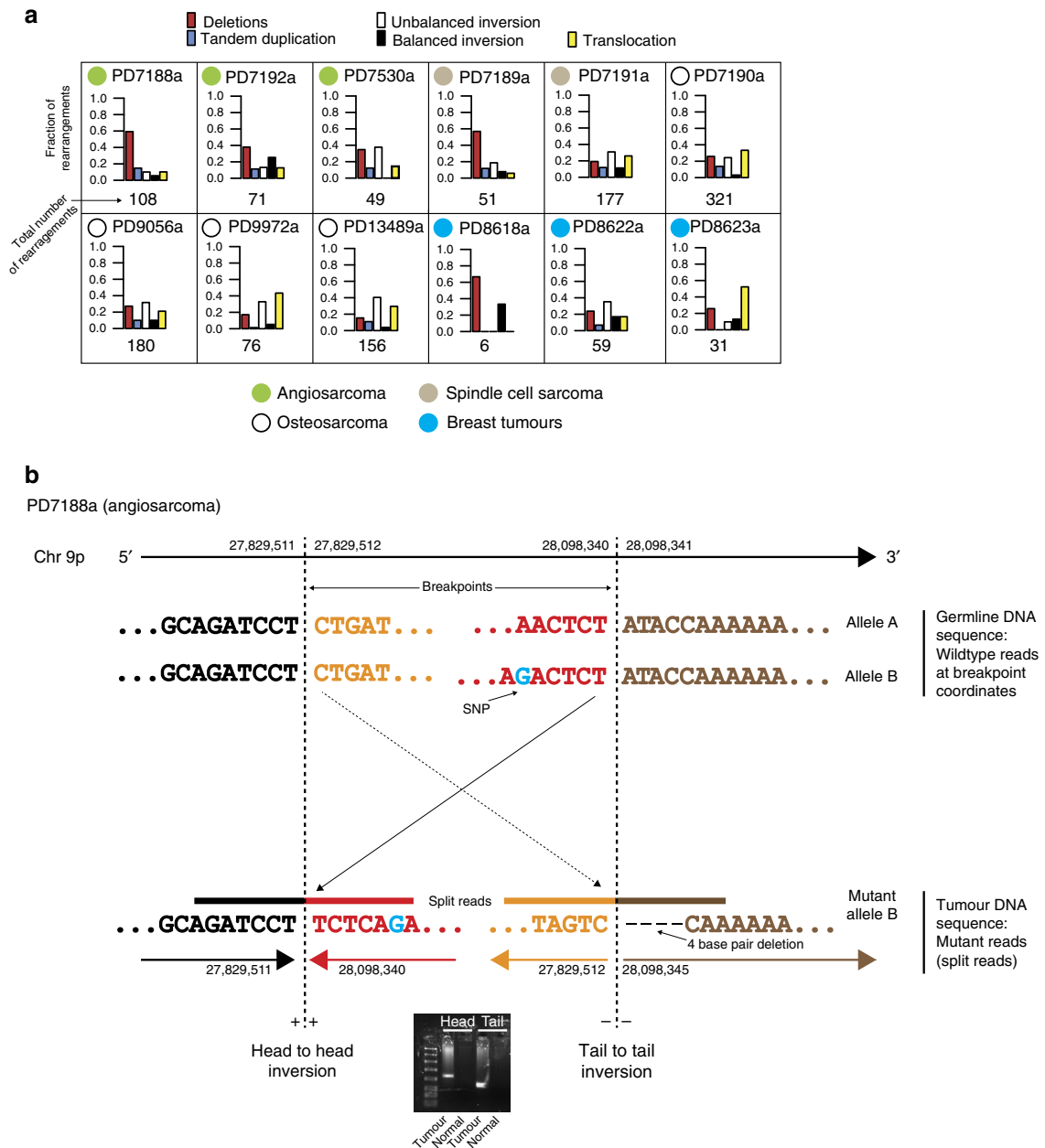
The relatively low number of mutations that we directly linked to ionizing radiation may seem surprising for such a well-known carcinogen. It is certainly considerably less than seen for cancers associated with tobacco, sunlight or aristolochic acid exposure<sup>10</sup>. This probably reflects the fact that although the attributable risk of such cancers is high, the absolute risk is relatively low. For example, >90% of angiosarcomas occurring after radiotherapy for primary breast cancer are attributable to radiation, but only one in a thousand women receiving such radiotherapy will develop angiosarcomas<sup>37</sup>, with a latency of many years. This suggests that although ionizing radiation clearly pushes bystander cells in the radiotherapy field towards cancer, the absolute burden of radiation-induced mutations per cell would not be high and additional driver mutations would be required.

**Figure 1 | Indels in radiation-associated tumours.** (a) Indel/substitution ratio. Shown is the indel/substitution ratio for each tumour. The ratio was significantly increased in radiation-associated second malignancies (0.0003, linear mixed effects model). Each dot represents a tumour. Different colours represent different tumour types (see legend, top right). Boxplots: vertical line – median; whiskers – minimum and maximum without outliers. (b) Deletion/insertion ratio. Shown is the deletion/insertion ratio of every tumour. Deletions were significantly (\*) enriched in radiation-associated second malignancies ( $P < 2.2 \times 10^{-16}$ , linear mixed effects model) and in breast tumours with germline *BRCA1* or *BRCA2* deficiency. Symbols, boxplots as per a. (c) Clonal versus subclonal indels in radiation-associated second malignancies. Shown are the absolute clonal (early) and subclonal (late) indel burdens of each tumour, by indel type. Amongst clonal indels, deletions were significantly enriched. *P*-values refer to the comparison of proportion of deletions/other indels in clonal versus subclonal indels (Fisher's exact test). (d) Indel likelihood across the genome. Shown is the probability of deletion or insertion to occur (vertical axis) across different regions of the genome (horizontal axis). The probability was modelled on the basis of associations between indels and genomic properties (see Methods). Chromosome 14 is shown as a representative chromosome. Radiation-associated indels were compared to indels of 35 non-radiation-associated osteosarcomas. Radiation-associated deletions, but not insertions, followed a more uniform distribution across the genome than in radiation-naïve samples. (e) Distribution of indels in relation to genomic features. Comparison of the mutation density of radiation versus non-radiation indels in relation to genomic features. X axis: ratio of mutation density of non-radiation-associated indels or radiation-associated indels over background density. Y axis: genomic feature. The distribution of insertions in both radiation-associated and radiation-naïve tumours correlated with several genomic features, with few significant differences (asterisk) between the two. In contrast, the distribution of deletions in radiation-induced cancers, but not in radiation-naïve tumours, showed little variability and resembled the background distribution more closely. Thus, significant differences (asterisk) were seen in the deletion density in relation to genomic features comparing radiation-associated and radiation-naïve tumours. *P*-values are detailed in Supplementary Data 4.

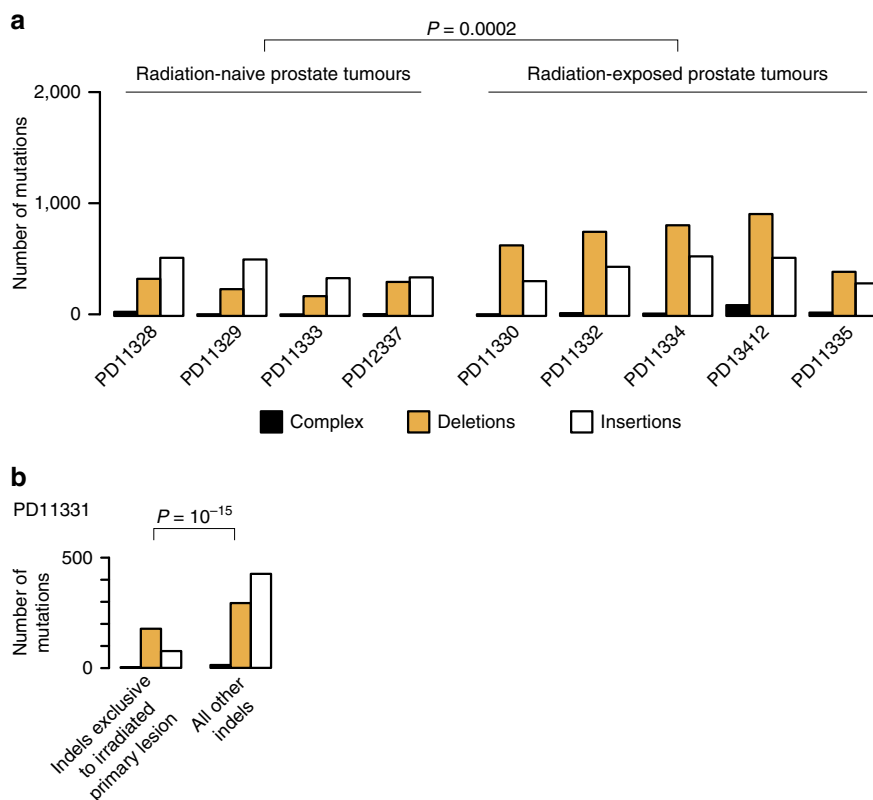
**Table 1 | Survey of balanced inversions in different tumour types.**

	Tumours with at least one balanced inversion	Number of tumours screened	Overall number of balanced inversions
Primary breast tumours	39	251	59
Primary osteosarcoma	4	35	7
<i>BRCA1/2</i> -deficient breast tumours	19	33	46
Radiation-associated second malignancies	11	12	52

Rearrangement catalogues of tumours were searched informatically for the presence of balanced inversions. The basic principle of the search was to find pairs of head-to-head and tail-to-tail inversions in which the breakpoint coordinates overlap at both ends. Compared with all 286 primary tumours, balanced inversions were significantly enriched in radiation-associated second malignancies ( $P = 2 \times 10^{-16}$ , generalized linear model) and also in *BRCA1* or *BRCA1* deficient breast tumours ( $P = 2 \times 10^{-16}$ , generalized linear model). Further, compared to *BRCA1* or *BRCA1* deficient breast tumours balanced inversions were significantly enriched in radiation-associated second malignancies ( $P = 0.0006$ , generalized linear model).



**Figure 2 | Balanced inversions in radiation-associated tumours. (a)** Overview of rearrangements. Tumours exhibited tumour-type specific features. Balanced inversions (black bars) were found in every tumour, except PD7530a. **(b)** Example of a balanced inversion in PD7188a. A 0.9 Mb inversion. The inversion was validated by PCR across the breakpoint (gel image) and by split reads. Note that the split reads carried a heterozygous SNP at the head end.



**Figure 3 | Indels in prostate tumours.** (a) Indels in radiation-naïve versus radiation-exposed prostate tumours. Shown is the indel burden, by indel subtype, found in radiation-naïve and in radiation-exposed prostate tumours. In radiation-exposed tumours radiotherapy had been administered to the primary tumour before formation of metastases. Deletions were significantly enriched in radiation-exposed tumours ( $P=0.0002$ , generalized linear model). Note that radiation-associated tumours with confounding BRCA1 or BRCA2 deficiency were excluded from the statistical analysis (Cases PD13412 and PD11335). (b) Indels in tumours from a patient whose primary lesion was treated with ionizing radiation after formation of metastases. Shown are indels that were found exclusively in the primary lesion and indels found in all other lesions. Deletions were significantly enriched amongst indels exclusive to the primary lesion. Comparison by Fisher's exact test, of the ratio of deletions over other indels.

## Methods

**Patient samples.** Informed consent was obtained from all subjects and ethical approval obtained from Cambridgeshire 2 Research Ethics Service (reference 09/H0308/165). Collection and use of patient samples were approved by the appropriate institutional review board of each Institution.

**Whole-genome sequencing.** DNA was extracted from 12 radiation-associated tumours and subjected to whole-genome sequencing, along with normal tissue derived from the same individuals. All tumour samples had been freshly frozen and were reviewed by reference pathologists. DNA extraction and preparation followed standard methods as previously described<sup>38</sup>. Reads were aligned to the reference human genome (NCBI37) by using BWA on default settings<sup>39</sup>. Reads which were unmapped or PCR-derived duplicates were excluded from the analysis. The average coverage of tumours was at least  $40\times$  and of normal DNA  $30\times$ , as per standard set by the International Cancer Genome Consortium.

**Variant detection.** The CaVEMan (cancer variants through expectation maximization) algorithm was used to call single-nucleotide substitutions (github.com/cancerit/CaVEMan). To call insertions and deletions, we used split-read mapping implemented as a modification of the Pindel algorithm<sup>38</sup>. To call rearrangements we applied the BRASS (breakpoint via assembly) algorithm, which identifies rearrangements by grouping discordant read pairs that point to the same breakpoint event (github.com/cancerit/BRASS). Post-processing filters were applied to the output to improve specificity. Copy-number data were derived from whole-genome reads using the ASCAT (version 2.2) algorithm<sup>40</sup>. Mutations were annotated to Ensembl version 58.

**Variant validation.** The precision of indels and substitutions presented here was assessed by manual inspection of 100 randomly selected substitutions and was found to be at least of 90% across the 12 radiation-associated tumours. This precision of coding indels and substitutions was confirmed by re-sequencing through whole-exome sequencing. Structural rearrangements were validated by defining exact break points through local reassembly, as implemented in BRASS.

Only rearrangements that could be validated have been included in this report (listed individually in Supplementary Data 3).

**Screen/validation of balanced inversions.** Rearrangement catalogues were screened for the presence of balanced inversions by means of a bespoke PERL script. Pairs of rearrangement calls were sought that were inversions in opposite directions with overlapping ranges of upper and lower break points. The search was directly performed on output from the Brass algorithm with the following post-processing filters: read count supporting the break point of greater than five reads and size of inversion greater than 2,500 base pairs unless the read count supporting the break point was greater than ten reads in which case no size threshold was applied. This post-processing strategy removes inversion artefacts, which are small and generally have a read count supporting the break point of less than five reads, without excluding small, high confidence inversions (defined as break points supported by at least ten reads). The precision of balanced inversion calls yielded by this search were assessed in the 12 radiation-associated tumours. In all but one balanced inversion, both rearrangements defining the inversion could be validated by algorithmic local reassembly or manual split-read mapping. In addition, a proportion of balanced inversions in 20/52 was subjected to PCR across the breakpoint in stock DNA from tumour and normal tissue. These inversions were all confirmed to be genuine and somatic (Supplementary Fig. 3).

**Germline variants.** Germline point mutations in *TP53*, *BRCA1* and *BRCA2* were searched for in catalogues of germline indels and substitutions, as determined by the point mutation variant calling algorithms employed here. Putative mutations were compared against publicly available catalogues of pathogenic germline mutations in these genes (www.iarc.fr).

**Extraction of substitution signatures.** Substitution signatures were extracted by using non-negative matrix factorization, as previously described<sup>10</sup>.

**Subclonality analyses.** Subclonal tumour cell populations that exist within tumours can be screened for by searching for non-heterozygous mutations in mutation catalogues, as previously described, using a Dirichlet process<sup>41</sup>. This was applied to substitution and indel catalogues of the 12 radiation-associated whole genomes. However, with indels there is a concern that mutant read frequencies may be underestimated for larger indels, as reads containing larger indels may be less amenable to mapping. To overcome this bias, the indel mutant read frequencies were corrected by extracting unmapped reads (split reads) from sequencing reads. The Dirichlet process was then applied to indel catalogues with, and without, correction. The results in terms of number of subclonal peaks were indistinguishable whether corrected or uncorrected indel catalogues were analysed. Three of the twelve genomes screened for subclones contained subclones which corresponded to subclones defined by substitutions in these tumours. Thus, the indel-defined subclones were considered genuine. Indels were subdivided into clonal (peak of mutation copy number  $\sim 1$ ) or subclonal (peak of mutation copy number  $< 1$ ).

**Non-radiation tumours.** A total of 319 tumours, 284 breast cancers and 35 osteosarcomas were included for comparison in analyses. These were spontaneous (primary), non-radiation-associated tumours. These tumours were sequenced to  $\sim 40\times$  or more, along with normal tissue DNA from the same patients. These tumours were prepared, sequenced, analysed by the same pipeline as the 12 radiation-associated tumours, including use of the same algorithms. The osteosarcoma cases were a series of paediatric and adult tumours (sequencing data published in the European Genome-phenome Archive, accession EGAD00001000147). The breast tumours were comprised of oestrogen receptor positive and negative tumours<sup>42</sup>. For the purposes of this analysis they were subdivided into spontaneous cases ( $n = 251$ ) and those associated with pathogenic germline *BRCA1* or *BRCA2* mutation ( $n = 33$ ). No control primary angiosarcoma and spindle cell sarcomas were available for inclusion in our analyses.

**Association of mutation density with genomic features.** The genomic properties listed in Supplementary Table 3 were calculated at every variant position, and, for comparison, at 100,000 random positions sampled uniformly from the callable regions of hg19. Only chromosomes 1–22 and X were considered. To test for differences in the genomic properties of variants in radiation-induced versus non-radiation-induced tumours, we used a two-proportion z-test for the binary variables, a t-test for the other quantitative variables (large sample size justifies central limit theorem), and a  $\chi^2$ -test for the categorical chromatin variable. A Benjamini-Yekutieli correction was applied to the raw P-values to account for multiple testing in the presence of likely correlation between these properties. Genomic properties are considered significantly different between radiation and non-radiation samples if the adjusted q-value is  $< 0.01$  and there is at least a 5% difference in magnitude between the two group means.

**Other statistical analyses.** To assess whether radiation-associated tumours harbour significantly more indels relative to substitutions and more deletions relative to insertions, a mixed linear effects model was implemented using the R package lme4. After incorporating as fixed effects type of mutation (substitution, deletion and insertion) and group of tumour, interactions between tumour group and type of mutation were assessed. For comparison of indel size distribution underlying the clustering in Supplementary Fig. 2a, the statistic of the Kolmogorov-Smirnov test was used (command in R: `ks.test(x,y)$statistic`). Unless indicated, R was used for calculations.

**Data availability.** Sequencing data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>), which is hosted by the European Bioinformatics Institute; accession numbers EGAS00001000138; EGAS00001000147; EGAS00001000195.

## References

- Leuraud, K. *et al.* Ionising radiation and risk of death from leukaemia and lymphoma in radiation-monitored workers (INWORKS): an international cohort study. *Lancet Haematol.* **2**, e276–e281 (2015).
- March, H. C. Leukemia in radiologists. *Radiology* **43**, 275–278 (1944).
- Travis, L. B. *et al.* Second malignant neoplasms and cardiovascular disease following radiotherapy. *J. Natl Cancer Inst.* **104**, 357–370 (2012).
- Preston, D. L. *et al.* Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. *Radiat. Res.* **158**, 220–235 (2002).
- Ronckers, C. M., Doody, M. M., Lonstein, J. E., Stovall, M. & Land, C. E. Multiple diagnostic X-rays for spine deformities and risk of breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **17**, 605–613 (2008).
- Cahan, W. G. *et al.* Sarcoma arising in irradiated bone; report of 11 cases. *Cancer* **1**, 3–29 (1948).
- Allan, J. M. & Travis, L. B. Mechanisms of therapy-related carcinogenesis. *Nat. Rev. Cancer* **5**, 943–955 (2005).
- Meier, B. *et al.* *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* **24**, 1624–1636 (2014).
- Flibotte, S. *et al.* Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics* **185**, 431–441 (2010).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Pleasant, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Poon, S. L. *et al.* Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci. Transl. Med.* **5**, 197ra101 (2013).
- Ravanat, J. L. *et al.* Radiation-mediated formation of complex damage to DNA: a chemical aspect overview. *Br. J. Radiol.* **87**, 20130715 (2014).
- Adewoye, A. B., Lindsay, S. J., Dubrova, Y. E. & Hurles, M. E. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nat. Commun.* **6**, 6684 (2015).
- Belfield, E. J. *et al.* Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Res.* **22**, 1306–1315 (2012).
- St Charles, J. *et al.* High-resolution genome-wide analysis of irradiated (UV and gamma-rays) diploid yeast cells reveals a high frequency of genomic loss of heterozygosity (LOH) events. *Genetics* **190**, 1267–1284 (2012).
- Ray, F. A. *et al.* Directional genomic hybridization: inversions as a potential biosimeter for retrospective radiation exposure. *Radiat. Environ. Biophys.* **53**, 255–263 (2014).
- Mitchell, C. R. *et al.* Stable intrachromosomal biomarkers of past exposure to densely ionizing radiation in several chromosomes of exposed individuals. *Radiat. Res.* **162**, 257–263 (2004).
- Natarajan, A. T., Boei, J. J., Vermeulen, S. & Balajee, A. S. Frequencies of X-ray induced pericentric inversions and centric rings in human blood lymphocytes detected by FISH using chromosome arm specific DNA libraries. *Mutat. Res.* **372**, 1–7 (1996).
- Saberi, A., Salari, E. & Latifi, S. M. Cytogenetic analysis in lymphocytes from radiation workers exposed to low level of ionizing radiation in radiotherapy, CT-scan and angiocardiology units. *Mutat. Res.* **750**, 92–95 (2013).
- Hande, M. P. *et al.* Past exposure to densely ionizing radiation leaves a unique permanent signature in the genome. *Am. J. Hum. Genet.* **72**, 1162–1170 (2003).
- Agrawala, P. K., Adhikari, J. S. & Chaudhury, N. K. Lymphocyte chromosomal aberration assay in radiation biodosimetry. *J. Pharm. Bioallied Sci.* **2**, 197–201 (2010).
- Gonin-Laurent, N. *et al.* Specific TP53 mutation pattern in radiation-induced sarcomas. *Carcinogenesis* **27**, 1266–1272 (2006).
- Gonin-Laurent, N. *et al.* RB1 and TP53 pathways in radiation-induced sarcomas. *Oncogene* **26**, 6106–6112 (2007).
- Nakanishi, H. *et al.* Mutation of the p53 gene in postradiation sarcoma. *Lab. Invest.* **78**, 727–733 (1998).
- Hadj-Hamou, N. S. *et al.* A transcriptome signature distinguished sporadic from postradiotherapy radiation-induced sarcomas. *Carcinogenesis* **32**, 929–934 (2011).
- Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
- Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
- Schuster-Bockler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
- Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
- Mehta, A. & Haber, J. E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **6**, a016428 (2014).
- Weinfeld, M. & Soderlind, K. J. 32P-postlabeling detection of radiation-induced DNA damage: identification and estimation of thymine glycols and phosphoglycolate termini. *Biochemistry* **30**, 1091–1097 (1991).
- Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
- Baca, S. C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
- Abbott, R. & Palmieri, C. Angiosarcoma of the breast following surgery and radiotherapy for breast cancer. *Nat. Clin. Pract. Oncol.* **5**, 727–736 (2008).
- Behjati, S. *et al.* Recurrent PTPRB and PLAG1 mutations in angiosarcoma. *Nat. Genet.* **46**, 376–379 (2014).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).



40. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).
41. Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* **5**, 2997 (2014).
42. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

### Acknowledgements

This work was supported by funding from the Wellcome Trust (grant reference 077012/Z/05/Z), Skeletal Cancer Action Trust, Rosetrees Trust UK, Bone Cancer Research Trust, the RNOH NHS Trust, the National Institute for Health Research Health Protection Research Unit in Chemical and Radiation Hazards and Threats at Newcastle University in partnership with Public Health England. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England. Tissue was obtained from the RNOH Musculoskeletal Research Programme and Biobank, co-ordinated by Mrs Deidre Brooking and Mrs Ru Grinnell, Biobank staff, RNOH. Support was provided to AMF by the National Institute for Health Research, UCLH Biomedical Research Centre, and the CRUK UCL Experimental Cancer Centre. S.N.Z. and S.B. are personally funded through Wellcome Trust Intermediate Clinical Research Fellowships, P.J.C. through a Wellcome Trust Senior Clinical Research Fellowship. We are grateful to the patients for participating in this research and to the clinicians and support staff involved in their care.

### Author contributions

S.B. and G.G. performed analyses of sequence data. D.C.W. and N.D.R. performed statistical analyses. P.S.T., M.R., H.D., S.N-Z contributed data and to data analysis. S.L.C. contributed to rearrangements analyses. P.V.L performed copy-number analysis. L.B.A. analysed substitution signatures. C.H. and C.L. performed technical investigations.

K.M.R., L.S., A.M., D.J., and R.S. contributed informatics tools. A.B. and J.W.T. co-ordinated informatics analyses. B.K. co-ordinated sample curation. P.A.F., A.S., C.B. and U.M. contributed to discussion. M.J., N.P., R.T., M.F.A., G.S.B., A.R. and A.M.F. curated samples, clinical data, and/or provided clinical expertise. A.M.F., M.R.S. and P.J.C. directed the research. S.B. and P.J.C. wrote the manuscript, with contributions from A.M.F., D.C.W., G.G. and G.S.B.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Behjati, S. *et al.* Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**:12605 doi: 10.1038/ncomms12605 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

### ICGC Prostate Group

Colin S. Cooper<sup>15,16</sup>, Rosalind A. Eeles<sup>16</sup>, Douglas Easton<sup>17</sup>, Christopher Foster<sup>18</sup>, David E. Neal<sup>19,20</sup>, Daniel S. Brewer<sup>15,16,21</sup>, Freddie Hamdy<sup>22</sup>, Yong-Jie Lu<sup>23</sup>, Andrew G. Lynch<sup>24</sup>, Charlie E. Massi<sup>19</sup>, Anthony Ng<sup>25</sup>, Hayley C. Whitaker<sup>19</sup>, Yongwei Yu<sup>26</sup>, Hongwei Zhang<sup>26</sup>, Elizabeth Bancroft<sup>16</sup>, Dan Berney<sup>23</sup>, Niedzica Camacho<sup>16</sup>, Cathy Corbishley<sup>27</sup>, Tokhir Dadaev<sup>16</sup>, Nening Dennis<sup>16</sup>, Tim Dudderidge<sup>28</sup>, Sandra Edwards<sup>16</sup>, Cyril Fisher<sup>28</sup>, Jilur Ghorri<sup>15</sup>, Vincent J. Gnanapragasam<sup>30</sup>, Christopher Greenman<sup>29</sup>, Steve Hawkins<sup>19</sup>, Steven Hazell<sup>28</sup>, Will Howat<sup>19</sup>, Katalin Karaszi<sup>22</sup>, Jonathan Kay<sup>19</sup>, Zsafia Kote-Jarai<sup>16</sup>, Barbara Kremeyer<sup>1</sup>, Pardeep Kumar<sup>28</sup>, Adam Lambert<sup>22</sup>, Daniel Leongamornlert<sup>16</sup>, Naomi Livni<sup>28</sup>, Hayley Luxton<sup>19</sup>, Lucy Matthews<sup>16</sup>, Erik Mayer<sup>16</sup>, Susan Merson<sup>16</sup>, David Nicol<sup>28</sup>, Christopher Ogden<sup>28</sup>, Sarah O'Meara<sup>1</sup>, Gill Pelvender<sup>31</sup>, Nimish C. Shah<sup>30</sup>, Simon Tavaré<sup>32</sup>, Sarah Thomas<sup>16</sup>, Alan Thompson<sup>28</sup>, Claire Verrill<sup>31</sup>, Anne Warren<sup>19</sup> & Jorge Zamora<sup>1</sup>

<sup>15</sup>Norwich Medical School and Department of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK; <sup>16</sup>Division of Genetics and Epidemiology, The Institute Of Cancer Research, London SW7 3RP, UK; <sup>17</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; <sup>18</sup>University of Liverpool and HCA Pathology Laboratories, London WC1E 6JA, UK; <sup>19</sup>Urological Research Laboratory, Cancer Research UK Cambridge Institute, Cambridge CB2 0RE, UK; <sup>20</sup>Department of Surgical Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; <sup>21</sup>The Genome Analysis Centre, Norwich NR4 7UH, UK; <sup>22</sup>The University of Oxford, Oxford OX1 2JD, UK; <sup>23</sup>Department of Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London EC1M 6BQ, UK.; <sup>24</sup>Statistics and Computational Biology Laboratory, Cancer Research UK Cambridge Institute, Cambridge CB2 0RE, UK; <sup>25</sup>The Chinese University of Hong Kong, Hong Kong, China; <sup>26</sup>Second Military Medical University, Shanghai 200433, China; <sup>27</sup>St George's Hospital, London SW17 0QT, UK; <sup>28</sup>Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK; <sup>29</sup>School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK; <sup>30</sup>Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK; <sup>31</sup>Oxford University Hospitals NHS Trust, John Radcliffe Hospital, Oxford OX3 9DU, UK; <sup>32</sup>Statistics and Computational Biology Laboratory, Cancer Research UK Cambridge Institute, Cambridge CB2 0RE, UK.