

SCIENTIFIC REPORTS



OPEN

Multisite *de novo* mutations in human offspring after paternal exposure to ionizing radiation

Manuel Holtgrewe^{1,4}, Alexej Knaus², Gabriele Hildebrand³, Jean-Tori Pantel³, Miguel Rodriguez de los Santos³, Kornelia Neveling⁵, Jakob Goldmann⁵, Max Schubach^{3,6}, Marten Jäger³, Marie Coutelier³, Stefan Mundlos³, Dieter Beule^{1,7}, Karl Sperling³ & Peter Michael Krawitz²

A genome-wide evaluation of the effects of ionizing radiation on mutation induction in the mouse germline has identified multisite *de novo* mutations (MSDNs) as marker for previous exposure. Here we present the results of a small pilot study of whole genome sequencing in offspring of soldiers who served in radar units on weapon systems that were emitting high-frequency radiation. We found cases of exceptionally high MSDN rates as well as an increased mean in our cohort: While a MSDN mutation is detected in average in 1 out of 5 offspring of unexposed controls, we observed 12 MSDNs in altogether 18 offspring, including a family with 6 MSDNs in 3 offspring. Moreover, we found two translocations, also resulting from neighboring mutations. Our findings indicate that MSDNs might be suited in principle for the assessment of DNA damage from ionizing radiation also in humans. However, as exact person-related dose values in risk groups are usually not available, the interpretation of MSDNs in single families would benefit from larger molecular epidemiologic studies on this new biomarker.

Improving our understanding of the effects of ionizing radiations on the human genome was one of the main motivations for the Department of Energy of the United States to initiate the Human Genome Project¹. As already noted 30 years ago, the most straightforward approach to looking for mutations is by determining the sequence of every nucleotide in a child's genome and then comparing this with the DNA sequence of the child's parents². Today, high-throughput sequencing technology enables us to analyze sequence variants at a genome-wide level and *de novo* rates for different mutational classes as well as their influencing factors have been investigated in several population-scale studies^{3–6}. However, the consequences of ionizing radiation for offspring of exposed humans have not yet been worked up with this technology and molecular epidemiological data on risk cohorts only exists on mini-satellites and chromosomal aneuploidies^{7–9}. While the results of these studies suggest a higher vulnerability to ionizing radiation during germ line maturation, a major limitation of these biomarkers is that the number of mutational events that can be measured per offspring is restricted. A recent study in mice pointed to multisite *de novo* mutations (MSDNs) as a potential biomarker that can be measured genome-wide and that might therefore be more sensitive to assess damage from high frequency radiation in single families^{10,11}. MSDNs are defined as two or more lesions within 20 bp and in offspring of male mice that were irradiated with 3 Gy, MSDNs occurred at a rate that is nine times higher than in controls (Fig. 1A)¹².

To our knowledge, we are the first to analyze MSDN rates in a cohort of human families with a potential exposure to high doses of ionizing radiation (IR). We analyzed one trio where the father was treated for Non Hodgkin lymphoma involving radiotherapy of the inguinal lymph nodes. Although the testes had been shielded during fractionated irradiation, the total dosage on the germline was estimated to be up to 2 Gy. Furthermore, we recruited 12 families with altogether 18 offspring, in which the fathers served as radar soldiers in the German military in the late 1950s to early 1980s. It was later found that some of the technical equipment they operated,

¹Berlin Institute of Health (BIH), Core Unit Bioinformatics, Berlin, 10178, Germany. ²Institute for Genomic Statistics and Bioinformatics, Rheinische Friedrich-Wilhelms Universität, Bonn, 53127, Germany. ³Charité – Universitätsmedizin Berlin, Institute of Medical and Human Genetics, 13353, Berlin, Germany. ⁴Charité – Universitätsmedizin Berlin, Berlin, 10117, Germany. ⁵Department for Human Genetics, Radboud University, Nijmegen, 6525, Netherlands. ⁶Berlin Institute of Health (BIH), JRG Computational Genome Biology, 10178, Berlin, Germany. ⁷Max Delbrück Center for Molecular Medicine, 13125, Berlin, Germany. Correspondence and requests for materials should be addressed to P.M.K. (email: pkrawitz@uni-bonn.de)

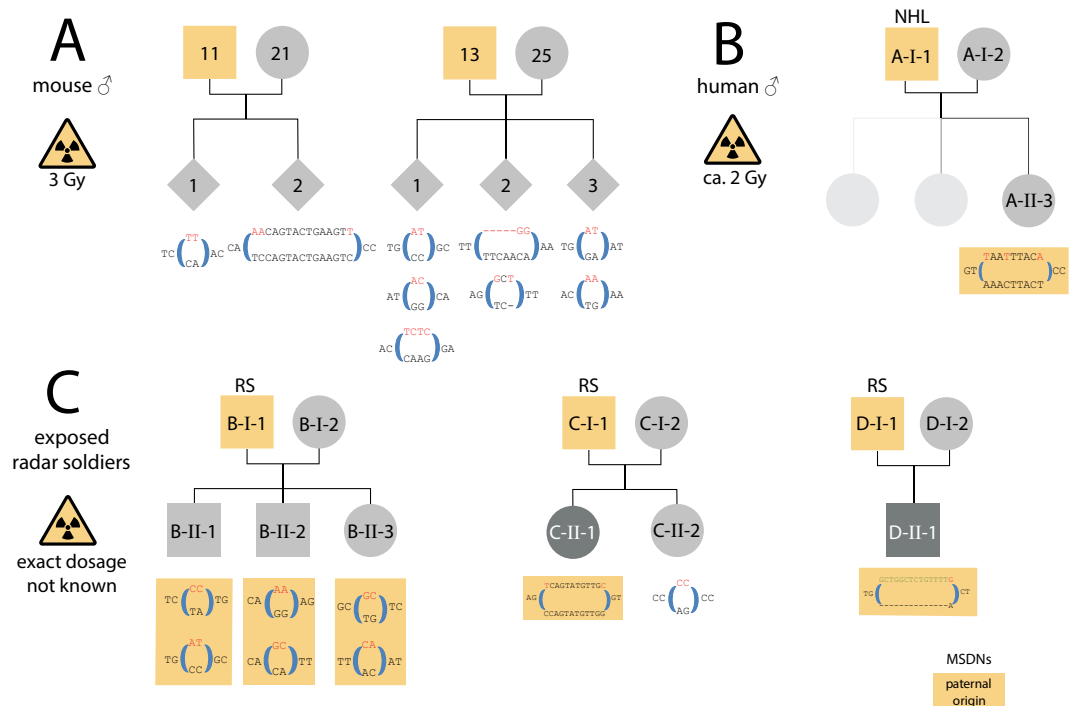


Figure 1. Multisite *de novo* mutations (MSDNs): (A) Offspring of male mice that were irradiated with 3 Gy showed a significantly increased rate of MSDNs¹. (B) The daughter of a Non Hodgkin Lymphoma (NHL) patient that was fathered after irradiation therapy had a MSDN of paternal origin. (C) Offspring of some radar soldiers (RS) that were potentially exposed to high frequency radiation during maintenance work, showed high rates of MSDNs. Depicted are families with at least one MSDN in at least one offspring. All 7 MSDNs for which the parental origin could be determined, are from the paternal germline (yellow shaded). Furthermore, a balanced (C-II-1) as well as an unbalanced translocation (D-II-1) could be identified that also originate from the paternal germline (see Fig. 2).

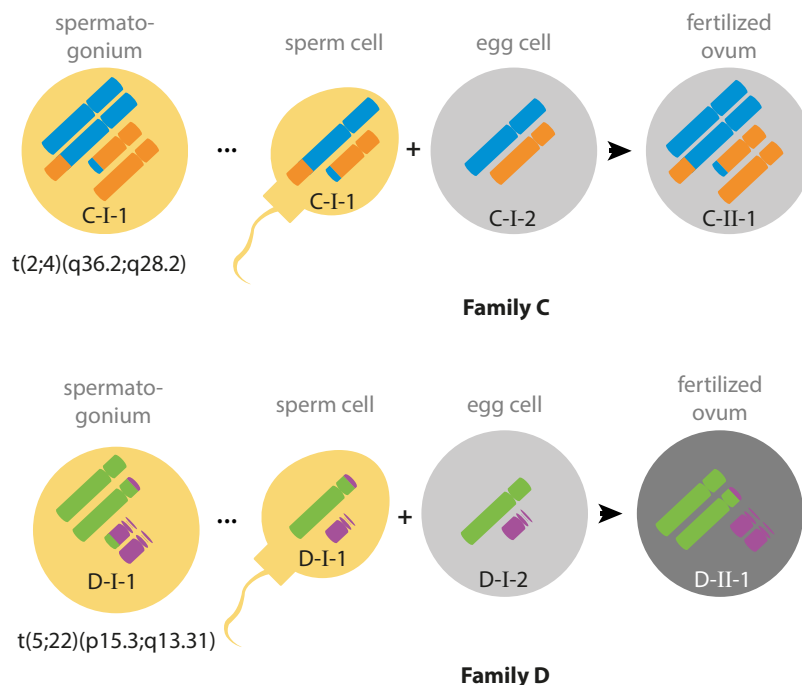


Figure 2. *de novo* chromosomal rearrangements in the paternal germline. Cri-du-chat syndrome observed in D-II-1, is explained by the *de novo* deletion of 5p, the molecular causes of the severe limb malformation in B-II-1 is yet under investigation (Supplemental Figs 7 and 8).

had not been optimally shielded and leaked high frequency radiation¹³. Weapon systems such as the HAWK, NIKE, SGR 103, and P15 exposed individuals to local doses of 150 mSv/h and even up to 400 mSv/h (SGR-103). Thus, during their entire duty at the radar units, these soldiers might have been exposed to high doses of IR and even though no exact doses can be determined in retrospective, different kinds of cancers have been recognized as occupational illnesses for this professional group.

Methods

Study cohort. In total, *de novo* mutations were analyzed in 18 offspring of radar soldiers with a potential exposure to ionizing radiation (IR) and 1 offspring of a Non-Hodgkin Lymphoma (NHL) patient who was subjected to radiotherapy. All experiments were carried out in accordance with the relevant guidelines and regulations. The ethics board of Charité-Universitätsmedizin Berlin approved this study, and informed consent was obtained from all participants. In addition, we analyzed the genomes of 28 offspring of parents without known exposure to IR. In the following, we also refer to the radar soldier cohort as “exposed” and their offspring as “cases” and the unexposed cohort as “controls”. In Fig. 1, we only depict pedigrees and MSDNs of four families that are discussed in more detail in the results section: Family A because of the good estimate for the dose, family B because of the extraordinary high rate of tandem *de novo* mutations, and families C and D because of the additional occurrence of chromosomal rearrangements. The additional three MSDNs of the exposed cohort that are not visualized in Fig. 1, as well as the MSDNs of the control cohort are listed in Table S3.

Sequencing and bioinformatics. Genomic DNA was obtained from blood or saliva samples and libraries were constructed according to the TruSeq PCR free or nano protocol depending on the amount of available DNA. After passing quality control, DNA fragments with an insert size between 340–570 bp were subjected to sequencing on a HiSeq X Ten platform with 150 bp paired-end protocol until minimum mean coverage of 30x per sample was reached. The sequence reads were aligned using BWA-MEM, duplicates were masked using Sambalster, and the resulting SAM files were converted to BAM and sorted using Samtools^{14–16}. SNVs and small indels were called using the UnifiedGenotyper (after realignment) and the HaplotypeCaller tools from GATK with default settings¹⁷.

SNVs and small indels were filtered to a high-confidence *de novo* candidate call sets in a fashion similar to the one described by Besenbacher *et al.*¹⁸ First, variants were filtered to those showing a *de novo* genotype pattern: heterozygous alternative in index and homozygous reference in the parents. Second, variants were excluded falling into the UCSC “simpleRepeat” or “repeat_masker” track and also those having a value below 1 in the UCSC track “wgEncodeCrgMapabilityAlign36mer”. Third, a minimum genotype quality of 50 was required, a coverage between 10 and 120 reads, and a fraction of alternative reads at the variant site between 0.2 and 0.8. Finally, we applied filters similar to Wong *et al.* to remove potential false positives caused by mapping errors⁵. For MSDNs that were subjected to Sanger validation and PacBio phasing, we dropped the criteria that no additional *de novo* calls within 1000 bp have been made, but we required detection by both calling tools, UnifiedGenotyper and HaplotypeCaller. The scripts that we used for filtering are also available upon request.

Phasing was performed on the high-confidence *de novo* SNV and small indel sets using the PhaseByTransmission and ReadBackedPhasing tools of GATK and then extracted using a similar approach as in Wong *et al.*⁵ As ReadBackedPhasing explores a number of haplotypes that is exponentially growing with the number of variants which leads to too high running times, we excluded regions with too many variants. For this, we split the genomes into windows of length 10 kbp and ignored them if they contained more than 100 variants. Of the *de novo* SNVs 34% could be phased, which is a ratio consistent with Wong *et al.*⁵.

As read-backed phasing with Illumina sequences is only possible over short distances, we generated long range PCR products for all MSDNs that could not be phased with Illumina data in the exposed cohort. PacBio reads were aligned with NGMLR version 0.2.2 and then postprocessed by the “bamChain” program provided by Pacific Biosciences¹⁹. Finally, the aligned reads were sorted using Samtools. Figure S5 shows an example of phasing a MSDN event in index B-II-1 using PacBio data. Two PacBio reads connect the allele with the MSDN in the child with the closest informative SNP in the father in a distance of roughly 4 kb.

We used IGV for viewing both Illumina and PacBio data and screened the MSDN candidates for nearby informative SNPs that could be used for assigning the *de novo* to maternal or paternal origin²⁰.

All MSDNs and phase informative SNPs in a distance below 500 bp were confirmed by Sanger sequencing of genomic DNA. Where the distance between the MSDN and the phase informative SNP exceeded 500 bp we sub-cloned the maternal and paternal alleles in the pT7T3U19 plasmid and sequenced the clones.

Results

In total, we subjected 19 offspring of potentially exposed fathers to whole genome sequencing (WGS) according to protocols that are similar to the large population studies that were recently conducted to assess the parental age influence on germline *de novo* mutations^{3–6}. We measured *de novo* mutation rates for single nucleotide variants (SNV) that are comparable to these studies and could also confirm the paternal and maternal age effect in our cohort (Supplemental Figs 1, 2 and 3). After accounting for age, we could not find significant differences in the total number of DNMs in the exposed and unexposed cohort.

In addition to isolated *de novo* SNVs, we also detected thirteen MSDNs in eleven offspring of seven fathers in our pilot study (Table 1, Fig. 1B,C). The *de novo* status has been confirmed for all MSDNs by Sanger sequencing and for eleven MSDNs we were also able to determine the parental origin by read-backed phasing: The MSDN in the NHL offspring as well as all 11 phased MSDNs in offspring of radar soldiers occurred in the paternal germline. 5 MSDNs were phased with informative heterozygous SNPs in more than 4 kb distance by long read sequencing data from the PacBio platform (see Figs S4 and S5). We also applied the same methods of MSDN detection to a control trio cohort of 28 offspring with limb malformations and could confirm only five MSDNs. Thus, the

Cohort	Subgroup	n	DNMs	Paternal	Maternal	NA	MSDN	P	M	NA
exposed	NHL	1	74	209	45	582	1	1	0	0
	Radar	18	762				12	11	0	1
control	Limb	28	1639	393	167	1079	5	3	0	2

Table 1. DNMs in exposed and unexposed cases and parental origin if ascertainable.

MSDN mutation rate per offspring in the potentially exposed cohort was 0.68 (13/19) compared to 0.18 (5/28) in the control.

Interestingly, there is a subclass of MSDNs with immediately adjoining nucleotide exchanges, also referred to as tandem mutations. We found two tandem mutations in the control cohort compared to a total of 1,639 *de novo* SNV calls (Table 1). This proportion is slightly smaller than the values of 0.3% and 0.4% that have previously been reported^{21,22}. In contrast, we found seven tandem mutations in the test cohort, six of them in three offspring of a single family (Fig. 1). Thus, in this family the relative tandem substitution rate is extraordinarily high, whereas the total number of *de novo* SNVs (B-II-1: 41, B-II-2: 44, B-II-3: 46) fits to the expectations for the parental ages (Supplemental Fig. 2). This ratio is astonishingly close to the observations of five tandems and 99 SNVs in Adewoye *et al.*¹¹. It is therefore an intriguing question whether not the total number of *de novo* SNVs is increased in human offspring after exposure to irradiation but only the fraction that is densely clustered and also more difficult to repair^{23,24}. *In silico* simulations do not only suggest that the number of lesions in a cluster is correlated positively with the linear energy transfer of the ionizing radiation, it is also a mutational class that might be especially difficult to repair when occurring post-meiotically^{24,25}.

Moreover, two offspring in families that also exhibit MSDNs, have *de novo* translocations. In C-II-1 the balanced translocation t(2;4)(q36.2;q28.2) was identified after multiple miscarriages and in D-II-1 cri-du-chat syndrome was molecularly confirmed by a deletion-duplication due to t(5;22)(p15.3;q13.31) (Fig. 2). In the WGS data, we found read-pairs that map to both chromosomes and that allowed us to identify the exact breakpoints. With informative SNPs on read-pairs that span the breakpoint, we were able to show that both translocations must have occurred in the paternal germlines (Figs S5 and S6).

Such translocations result from two closely neighboring DSBs on different chromosomes that are inappropriately repaired by non-homologous end joining. The “spontaneous” incidence of such reciprocal translocations is below one in a thousand pregnancies^{26,27}.

Conclusion

We analyzed the MSDNs in a cohort of offspring that were conceived while their fathers were at risk of exposure to high frequency radiation and observed rates that were higher than in a control cohort that was analyzed with the same protocols. Personal communications with the Inova study group confirmed that the event rate for paternal MSDNs in the general population is estimated to be 0.23 and the Poisson probability to observe 11 or more paternal MSDNs in 18 offspring with this rate parameter is below 0.005²⁸. The results of our small pilot study are certainly encouraging to analyze MSDN rates in larger case cohorts and we started recruiting for a follow up study (<https://www.igsb.uni-bonn.de/en/research/radar>). With highly accurate long-read sequencing technology it is especially intriguing to hypothesize that radiation induced mutations in mini-satellites in combinations MSDNs in offspring genomes might serve as a dosimeter for the cumulative radiation exposure of their parents.

References

- Watson, J. J. D. & Cook-Deegan, R. R. M. R. Origins of the Human Genome Project. *Faseb J.* **8**, 8–11 (1991).
- U.S. Congress OoTA. Technologies for Detecting Heritable Mutations in Human Beings. OTA-H-298 (1986)
- Jónsson, H. *et al.* Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017).
- Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471 (2012).
- Wong, W. S. W. *et al.* New observations on maternal age effect on germline *de novo* mutations. *Nat. Commun.* **7**, 10486 (2016).
- Kloosterman, W. P. *et al.* Characteristics of *de novo* structural changes in the human genome. *Genome Res.* **25**, 792–801 (2015).
- Dubrova, Y. E. *et al.* Human minisatellite mutation rate after the Chernobyl accident. *Nature* **380**, 683–686 (1996).
- Livshits, L. A. *et al.* Children of chernobyl cleanup workers do not show elevated rates of mutations in minisatellite alleles. *Radiat. Res.* **155**, 74–80 (2001).
- Sperling, K. *et al.* Significant increase in trisomy 21 in Berlin nine months after the Chernobyl reactor accident: Temporal correlation or causal relation? *BMJ* **309**, 158 (1994).
- Nakamura, N., Suyama, A., Noda, A. & Kodama, Y. Radiation effects on human heredity. *Annu. Rev. Genet.* **47**, 33–50 (2013).
- 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).
- Georgakilas, A. G., O'Neill, P. & Stewart, R. D. Induction and Repair of Clustered DNA Lesions: What Do We Know So Far? *Radiat. Res.* **180**, 100–109 (2013).
- König, W., Bericht der Expertenkommission zur Frage der Gefährdung durch Strahlung in früheren Radareinrichtungen der Bundeswehr und der NVA, Bundesamt für Strahlenschutz, 2. Juli (2003)
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Faust, G. G. & Hall, I. M. SAMBLASTER: Fast duplicate marking and structural variant read extraction. *In Bioinformatics* **30**, 2503–2505 (2014).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
- Besenbacher, S. *et al.* Novel variation and *de novo* mutation rates in population-wide *de novo* assembled Danish trios. *Nat. Commun.* **6** (2015).
- Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single molecule sequencing. bioRxiv 169557, <https://doi.org/10.1101/169557> (2017).

20. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24–26 (2011).
21. Chen, J. M., Cooper, D. N. & Férec, C. A new and more accurate estimate of the rate of concurrent tandem-base substitution mutations in the human germline: ~0.4% of the single-nucleotide substitution mutation rate. *Hum. Mutat.* **35**, 392–394 (2014).
22. Besenbacher, S. *et al.* Multi-nucleotide *de novo* Mutations in Humans. *PLoS Genet.* **12** (2016).
23. Horai, M. *et al.* Detection of *de novo* single nucleotide variants in offspring of atomic-bomb survivors close to the hypocenter by whole-genome sequencing. *J. Hum. Genet.* **63**, 357–363 (2018).
24. Frankenberg-Schwager, M. Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells. *Radiat. Environ. Biophys.* **29**, 273–292 (1990).
25. Eccles, L. J., O'Neill, P. & Lomax, M. E. Delayed repair of radiation induced clustered DNA damage: Friend or foe? *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **711**, 134–141 (2011).
26. Giardino, D. *et al.* *de novo* balanced chromosome rearrangements in prenatal diagnosis. *Prenat. Diagn.* **29**, 257–265 (2009).
27. Richardson, C. & Jasin, M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405**, 697–700 (2000).
28. Bernard, R. *Fundamentals of Biostatistics*, (7th edition). Thomson-Brooks/Cole, <https://doi.org/10.1002/bimj.4710350205> (2010).

Acknowledgements

We are thankful for the support of the Bund zur Unterstützung der Radargeschädigten (BzUR e.V.) and its members for their participation in the study. WGS costs were covered by a generous donation from Dr. Gisela Sperling. We thank the reviewers, Christian Gilissen, Wendy SW Wong, and John E Niederhuber for personal communications and fruitful discussion on the event rate of paternal MSDNs in the Inova study.

Author Contributions

M.H., M.J., M.S., J.G. performed the bioinformatics analysis, M.R., A.K. and G.H. performed Sanger validations, J-T.P. created the figures, M.C., A.K., and K.N. analyzed PacBio data. M.H., S.M., D.B., K.S., and P.M.K. designed the study and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33066-x>.

Competing Interests: The authors declare no competing interests.

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) 2018