



Published in final edited form as:

Science. 2021 May 14; 372(6543): . doi:10.1126/science.abg2538.

Radiation-related genomic profile of papillary thyroid cancer after the Chernobyl accident

Lindsay M. Morton^{#1,*}, Danielle M. Karyadi^{#2,*}, Chip Stewart^{#3,*}, Tetiana I. Bogdanova^{4,†}, Eric T. Dawson^{2,5,†}, Mia K. Steinberg^{6,†}, Jieqiong Dai⁶, Stephen W. Hartley², Sara J. Schonfeld¹, Joshua N. Sampson⁷, Yosi Maruvka³, Vidushi Kapoor⁶, Dale A. Ramsden⁸, Juan Carvajal-Garcia⁹, Charles M. Perou^{10,11}, Joel S. Parker¹¹, Marko Krznaric¹², Meredith Yeager⁶, Joseph F. Boland⁶, Amy Hutchinson⁶, Belynda D. Hicks⁶, Casey L. Dagnall⁶, Julie M. Gastier-Foster^{13,14}, Jay Bowen¹³, Olivia Lee², Mitchell J. Machiela¹⁵, Elizabeth K. Cahoon¹, Alina V. Brenner¹⁶, Kiyohiko Mabuchi¹, Vladimir Drozdovitch¹, Sergii Masiuk¹⁷, Mykola Chepurny¹⁷, Liudmyla Yu. Zurnadzy⁴, Maureen Hatch¹, Amy Berrington de Gonzalez¹, Gerry A. Thomas^{12,‡}, Mykola D. Tronko^{18,‡}, Gad Getz^{3,19,20,‡}, Stephen J. Chanock^{2,‡}

¹Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, United States

²Laboratory of Genetic Susceptibility, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, United States

³Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, United States

⁴Laboratory of Morphology of the Endocrine System, V.P.Komisarenko Institute of Endocrinology and Metabolism of the National Academy of Medical Sciences of Ukraine, Kyiv, 04114, Ukraine

Correspondence to: chanocks@mail.nih.gov, mortonli@mail.nih.gov.

Author contributions: The following authors conceptualized the study: L.M.M, D.M.K., C.S., T.I.B., E.T.D., A.V.B., A.B.d.G., G.A.T., M.D.T., G.G., and S.J.C. The following authors designed the study methodology: L.M.M, D.M.K., C.S., T.I.B., E.T.D., M.K.S., J.D., S.W.H., S.J.S., J.N.S., Y.M., V.K., D.A.R., J.C.G., C.M.P., J.S.P., M.Y., O.L., M.J.M., A.V.B., V.D., S.M., M.C., L.Y.Z., G.A.T., G.G., and S.J.C. The following authors analyzed and synthesized the data: L.M.M, D.M.K., C.S., E.T.D., M.K.S., J.D., S.W.H., S.J.S., Y.M., V.K., M.Y., O.L., M.J.M., V.D., A.B.d.G., G.A.T., G.G., and S.J.C. The following authors collected or generated the study data: C.S., T.I.B., E.T.D., M.K.S., J.D., M.Y., J.F.B., A.H., B.D.H., C.L.D., J.M.G.F., J.B., E.K.C., S.M., M.C., L.Y.Z., G.A.T., and M.D.T. The following authors contributed resources (study materials, patients, computing resources, or other analysis tools): T.I.B., M.K.S., J.D., M.K., A.V.B., K.M., V.D., M.H., A.B.d.G., G.A.T., M.D.T., G.G., and S.J.C. The following authors curated the research data: L.M.M, D.M.K., C.S., T.I.B., E.T.D., V.K., M.K., A.H., C.L.D., J.M.G.F., and J.B. The following authors wrote the original draft: L.M.M, D.M.K., C.S., E.T.D., M.K.S., J.D., S.W.H., S.J.S., M.Y., G.A.T., G.G., and S.J.C. The following authors visualized the data: L.M.M, D.M.K., C.S., T.I.B., E.T.D., M.K.S., J.D., V.K., O.L., and G.G. The following authors supervised or managed the research: L.M.M, D.M.K., C.S., G.A.T., M.D.T., G.G., and S.J.C. The following authors had responsibility for managing and coordinating the research activity: L.M.M, D.M.K., C.S., T.I.B., E.T.D., E.K.C., A.B.d.G., G.A.T., M.D.T., G.G., and S.J.C. The following authors acquired funding for the study: G.A.T., M.D.T., and S.J.C. All authors edited the final manuscript.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡These authors jointly directed this work.

Competing interests:

E.T.D. is an employee of Nvidia Corporation and owns stock in Nvidia, Illumina, and Pacific Biosciences. G.G. receives research funds from IBM and Pharmacyclics, and is an inventor on patent applications related to MuTect, ABSOLUTE, MutSig, MSMuTect, MSMuTect, MSIdetect, POLYSOLVER and TensorQTL. G.G. is a founder, consultant and holds privately held equity in Scorpion Therapeutics. All other authors declare no competing interests.

Data and materials availability: Analytic data are available in Data S1–S3. The Materials and Methods text specifies code that has been posted to GitHub and is archived on Zenodo (13). Raw molecular data are available from the Genomic Data Commons, accessed through the database of Genotypes and Phenotypes (dbGaP, accession phs001134; <https://www.ncbi.nlm.nih.gov/gap/>).

⁵Nvidia Corporation, Santa Clara, California, 95051, United States

⁶Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Bethesda, Maryland, 20892, United States

⁷Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, United States

⁸Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, United States

⁹Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, United States

¹⁰Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, United States

¹¹Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, United States

¹²Department of Surgery and Cancer, Imperial College London, Charing Cross Hospital, London, W6 8RF, United Kingdom

¹³Nationwide Children's Hospital, Biospecimen Core Resource, Columbus, Ohio, 43205, United States

¹⁴Departments of Pathology and Pediatrics, Ohio State University College of Medicine, Columbus, Ohio, 43210, United States

¹⁵Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, United States

¹⁶Radiation Effects Research Foundation, Hiroshima, 732-0815, Japan

¹⁷Radiological Protection Laboratory, Institute of Radiation Hygiene and Epidemiology, National Research Center for Radiation Medicine of the National Academy of Medical Sciences of Ukraine, Kyiv, 04050, Ukraine

¹⁸Department of Fundamental and Applied Problems of Endocrinology, V.P.Komisarenko Institute of Endocrinology and Metabolism of the National Academy of Medical Sciences of Ukraine, Kyiv, 04114, Ukraine

¹⁹Center for Cancer Research and Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, 02114, United States

²⁰Harvard Medical School, Boston, Massachusetts, 02115, United States

These authors contributed equally to this work.

Abstract

The 1986 Chernobyl nuclear power plant accident increased papillary thyroid cancer (PTC) incidence in surrounding regions, particularly for ¹³¹I-exposed children. We analyzed genomic, transcriptomic, and epigenomic characteristics of 440 PTCs from Ukraine (359 with estimated childhood ¹³¹I exposure and 81 unexposed children born after 1986). PTCs displayed radiation

dose-dependent enrichment of fusion drivers, nearly all in the mitogen-activated protein kinase pathway, and increases in small deletions and simple/balanced structural variants that were clonal and bore hallmarks of non-homologous end-joining repair. Radiation-related genomic alterations were more pronounced for those younger at exposure. Transcriptomic and epigenomic features were strongly associated with driver events but not radiation dose. Our results point to DNA double-strand breaks as early carcinogenic events that subsequently enable PTC growth following environmental radiation exposure.

One Sentence Summary:

Post-Chernobyl papillary thyroid cancers demonstrate radiation dose-dependent increases in clonal DNA double-strand breaks.

The accidental explosion in reactor 4 at the Chernobyl (Chornobyl in Ukrainian) nuclear power plant in April 1986 resulted in the exposure of millions of inhabitants of the surrounding areas of Ukraine, Belarus, and the Russian Federation to radioactive contaminants (1).

Epidemiologic and clinical research in the ensuing decades has demonstrated increased risk of papillary thyroid carcinoma (PTC) with increasing thyroid gland exposure to radioactive iodine (^{131}I) from fallout, which was deposited on pastures with grazing cows and ingested through milk and leafy greens, particularly during early childhood (2). Together with data from populations exposed to other types of radiation, compelling evidence indicates that PTC risk increases following childhood exposure to ionizing radiation, a recognized carcinogen (2–5).

Currently, there are no established molecular biomarkers for cancers induced by radiation, nor have there been large-scale analyses of the genomic landscape of human cancers occurring after a well-quantified radiation exposure. Classical cytogenetic studies have demonstrated radiation dose-associated increases in large chromosomal aberrations (such as inversions and translocations) that reflect DNA double-strand breaks and are the current standard for biodosimetry; however, these assays are typically performed in peripheral blood lymphocytes from individuals exposed to whole body irradiation and have not been directly linked to tumor characteristics (6, 7). Next-generation sequencing of 12 second primary tumors of various types that occurred within the field of previous therapeutic ionizing radiation suggested an excess of small deletions and balanced inversions (8), but radiation dose estimates were not available.

RNA sequencing (RNA-seq) analyses of 65 PTCs (mean age at diagnosis=24.7 years) occurring after the Chernobyl accident identified that higher doses were associated with an increased likelihood of gene fusion drivers (9). In a genomic landscape analysis of 496 primarily unexposed PTCs (mean age at diagnosis=46.8 years; 16 patients with known prior radiation exposure), The Cancer Genome Atlas (TCGA) reported a low density of somatic point mutations, especially for PTCs in younger patients, and a high frequency of activating somatic alterations in the mitogen-activated protein kinase (MAPK) pathway, including

point mutations in *BRAF* (61.7%) and *NRAS/HRAS/KRAS* (12.9%) as well as fusions with *RET* (6.8%), *BRAF* (2.7%) and other MAPK-related genes (5.0%) (10).

Here, we report a comprehensive characterization of the genomic, transcriptomic, and epigenomic profile of PTC as well as non-tumor thyroid tissue and/or blood for 440 individuals from Ukraine who developed PTC after the Chernobyl accident (mean age at PTC=28.0 years), affording an opportunity to investigate the contribution of environmental radiation to PTC characteristics. The study analyzed a collection of pre-treatment fresh frozen tumor tissues with pathological confirmation of first primary PTC by an international panel of experts through the Chernobyl Tissue Bank (CTB) (11, 12). Our study included 359 individuals with PTC with well-quantified ¹³¹I-exposure before adulthood (18 years of age; mean=7.3 years) and, as controls, 81 ¹³¹I-unexposed individuals with PTC born >9 months after the Chernobyl accident (all were born after March 1, 1987) (13).

RESULTS

Samples, Clinical Data, and Analytic Approach

Based on the availability of sufficient DNA and RNA extracted from CTB samples, we analyzed up to 440 individuals with whole genome sequencing (WGS) and/or mRNA-seq of pathologically confirmed tumor (374 both, 57 mRNA-seq only, 9 WGS only) (Fig. S1–2) (13). Matched normal tissue with WGS and/or mRNA-seq included non-tumor thyroid tissue and/or blood (233 both, 182 non-tumor tissue only, 16 blood only, 9 normal tissue not available). The genomic landscape characterization was augmented by single nucleotide polymorphism (SNP) microarray genotyping (Illumina Infinium HumanOmniExpress-24 array) and relative telomere length quantification on all samples, and DNA methylation profiling (Illumina Infinium MethylationEPIC array) and microRNA (miRNA)-seq for PTC and non-tumor thyroid tissue (Fig. S3). A total of 357 individuals had tumor sample data across all platforms.

The majority of individuals with PTC were female (n=335, 76.1%), resided in the Kiev (Kyiv in Ukrainian) region at the time of the accident (n=286, 65.0%), and were diagnosed during young adulthood (mean=28.0 years, range: 10.0–45.6); ¹³¹I-unexposed individuals with PTC were born at least 9 months after the accident and thus had a younger average age than the exposed individuals (mean: unexposed=20.7 years, exposed=29.7 years) (Table S1, Fig. S4). For ¹³¹I-exposed individuals, mean age at exposure was 7.3 years (range: in utero to 18.9 years) and mean time from the accident to PTC diagnosis was 22.4 years (range: 12.5–29.9). Radiation doses to the thyroid were reconstructed by an international team of dosimetry experts (14–16). For 53 individuals, doses were estimated using detailed information derived from individual direct thyroid radioactivity measurements taken within 8 weeks of the accident, with (n=49) or without (n=4) personal interviews regarding residential history and dietary patterns. For the remaining individuals, dose estimates were derived from direct measurements taken for other individuals who lived in the same residential area (n=249), neighboring area (n=9), or other areas (n=39), or based on dose estimates to the mother for individuals who were in utero at the time of the accident (n=9). Mean estimated radiation dose was 250 mGy (range: 11–8,800) (Fig. S4–S5).

Our primary analyses investigated the relationship between ^{131}I dose and 68 PTC molecular characteristics derived from a comprehensive genomic landscape analysis (Fig. 1) using multivariable linear, proportional odds, or logistic regression models adjusted for sex and age at PTC diagnosis (13). For associated variables (defined as $P < 7.4 \times 10^{-4}$ based on a Bonferroni correction for 68 tests), further analyses were conducted by specific molecular characteristics, as well as by age at PTC, age at exposure, and time since exposure (latency) because these factors influence radiation-related thyroid cancer risk (17). In addition, we conducted sensitivity analyses to assess whether the results were consistent when we restricted the population to ^{131}I -exposed individuals with lower radiation dose (< 500 mGy, resulting in $n=326$ with mean dose= 110 mGy) (18, 19).

Simple somatic variants

WGS analysis of tumor/normal pairs ($n=383$; mean sequencing depth, tumor= 89X , non-tumor thyroid tissue= 33X , blood= 33X ; Table S2) revealed a low burden of simple somatic variants (SSV) (mean= 0.27 nonsynonymous mutations per Mb) (Fig. 1, Fig. S6–S7), which was lower than in older TCGA PTC cases (0.41 nonsynonymous mutations per Mb) (10) and comparable to mutationally quiet tumors typically reported for pediatric cancers (20). A total of $318,956$ SSVs were identified, the majority (93.3%) of which were single nucleotide variants (SNVs) ($n=297,512$ in 383 tumors; mean per tumor= 776.8), whereas small insertions and deletions (indels) were less common (insertions: $n=5,842$, 1.8%, mean= 37.2 ; deletions: $n=14,231$, 4.5%, mean= 15.3), and doublet and triplet base substitutions were rare (dinucleotide polymorphism [DNP] or doublet: $n=1,351$, 0.4%, mean= 3.5 ; trinucleotide polymorphism [TNP] or triplet: $n=20$, 0.006%) (Table S3). Among the $3,886$ coding mutations (1.2% of total; $0.35/\text{Mb}$), most were nonsynonymous ($3,023/3,886=77.8\%$). Approximately one-third of mutations ($n=114,898$, 36.0%) were clonal (cancer cell fraction 0.9), regardless of mutation type (SNV= 35.9% , insertions= 36.7% , deletions= 38.0%) (Table S3, Fig. S8) (13).

In multivariable analyses restricted to $n=356$ samples with both high tumor and normal tissue purity (Fig. 1, Fig. S9) (13), increased radiation dose was associated with an increase in small deletions ($P=8.0 \times 10^{-9}$) as well as the deletion:SNV ratio ($P=4.9 \times 10^{-21}$), but not other SSV types (Fig. 2, Table S4). In addition, we observed the expected increase in the burden of SNVs ($P=3.2 \times 10^{-6}$), doublet mutations ($P=2.7 \times 10^{-5}$), insertions ($P=1.5 \times 10^{-6}$), and deletions ($P=7.4 \times 10^{-16}$) with increasing age at PTC diagnosis (Table S4) (10). Few of these mutations were clustered (22 clusters [>2 mutations within 150 base pairs (bp)] in 18 cases; 83 clusters [>2 mutations within 1 kb] in 36 cases) and were not associated with radiation dose ($P > 0.3$). In an analysis of the frequency and types of microsatellite indels in the tumors, detected using MSMuTect (21), all tumors were microsatellite stable (mean [range] per tumor, insertions= 1.8 [0–7], deletions= 7.3 [0–24]), and radiation dose was not significantly associated with the number of microsatellite insertions or deletions (Table S4).

An investigation of mutational processes in PTC was conducted using SigProfiler to determine both single base substitution (SBS) and small indel (ID) mutational signatures (20, 22). Comparing the PTC mutations with known signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC v3, <https://cancer.sanger.ac.uk/cosmic/signatures>)

(20), the majority of the SBS signatures (69.9%) were attributable to clock-like signatures (SBS1=9.8%, SBS5=60.2%), with smaller fractions due to APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) cytidine deaminase DNA-editing activity (SBS2=6.2%, SBS13=6.4%), damage from reactive oxygen species (SBS18=0.9%), and two signatures of unknown etiology (SBS8=15.1%, SBS23=1.6%) (mean cosine similarity between actual mutations and attributed patterns=0.94) (Fig. 1, Table S5, Fig. S10–S12). In multivariable analyses, no SBS signatures were significantly associated with radiation dose, whereas increased age at PTC diagnosis was associated with an increase in clock-like SBS mutations (PSBS1= 1.9×10^{-7} ; PSBS5= 6.8×10^{-17}) as well as SBS8 ($P=4.3 \times 10^{-11}$) and SBS18 ($P=2.0 \times 10^{-8}$) (Table S4).

The majority (54.0%) of indels were attributed to clock-like signatures (ID1=14.2%, ID5=39.8%), 21.2% to ID3 (tobacco smoking, which is not a major risk factor for PTC), 19.0% to repair of DNA double-strand breaks by end-joining mechanisms (ID6=3.3%, ID8=15.8%), and 5.8% to ID4 (unknown etiology) (mean cosine similarity=0.77) (Fig. 1, Table S5, Fig. S10–S12). In multivariable analyses, radiation dose was strongly associated with end-joining-related indel mutational patterns ($P=1.5 \times 10^{-10}$), particularly ID8 ($P=7.3 \times 10^{-9}$), and more weakly with the clock signature ID5 ($P=1.3 \times 10^{-4}$) (Fig. 2, Table S4). In comparison, increased age at PTC diagnosis was associated with significantly increased numbers of ID3 ($P=1.9 \times 10^{-7}$), ID5 ($P=1.9 \times 10^{-9}$), and ID8 ($P=6.6 \times 10^{-4}$) mutational patterns (Table S4). *De novo* signature extraction did not reveal a novel signature related to environmental exposure to ionizing radiation but identified 4 SBS (mean cosine similarity=0.96) and 2 ID (mean cosine similarity=0.83) signatures highly correlated with COSMIC signatures described above (Table S4, Table S6, Fig. S10–S12). Similarly, no novel signature was identified when we restricted the analysis to PTCs in individuals who received 200 mGy (Table S5).

Structural variation

Overall, 479 structural variants (SV) were identified in 356 tumors; approximately one-quarter of SVs ($n=132$, 27.6%) were simple/balanced events (balanced interchromosomal translocations and inversions), one-half ($n=253$, 52.8%) were simple/unbalanced (such as deletions, unbalanced interchromosomal translocations), and the remaining ($n=94$, 19.6%) were complex (such as >2 breaks repaired together in a cluster) (Fig. S13–14, Table S6) (13). Approximately one-third of tumors ($n=113$, 31.7%) had no SV, one-third ($n=126$, 35.4%) had one SV, and the remaining third had two ($n=61$, 17.1%) or more ($n=56$, 15.7%) SVs. Two tumors had >10 SV events (Fig. S15), one of which was the only tumor with evidence for chromothripsis (age at PTC=30.2 years, age at exposure=1.3 years, dose=1000 mGy).

Multivariable analyses ($n=354$, excluding the 2 outliers) demonstrated that increasing radiation dose was significantly associated with increased SV count ($P=1.4 \times 10^{-8}$), particularly simple/balanced SVs ($P=1.2 \times 10^{-14}$) but not those classified as complex ($P=0.52$) or simple/unbalanced ($P=5.6 \times 10^{-3}$) (Fig. 3A, Table S4). Increasing radiation dose also was not associated with occurrence of chromoplexy ($P=0.70$) (Table S4), which was

identified in 19 tumors (n=15 single event, n=4 two events) (13, 23), nearly all unexposed or in the lower dose groups ($n_{\text{unexposed}}=7$, $n_{1-99 \text{ mGy}}=8$, $n_{100-199 \text{ mGy}}=2$, $n_{\geq 500 \text{ mGy}}=2$).

Somatic copy number alteration

A total of 40.3% (n=143/355) of the tumors evaluated for somatic copy number alterations (SCNA) had one (n=96, 27.0%) or more (n=47, 13.2%) such events (Fig. S16) (13). Four tumors had ≥ 20 SCNAs each (Fig. S17): the tumor with chromothripsis and three additional tumors (age at PTC=19.0–29.3 years, age at exposure=1.3–2.3 years, dose=125–175 mGy). Those three tumors predominantly had gains or copy neutral loss of heterozygosity (CNLOH) and were the only tumors with ploidy >2.5 , with one of the three displaying extensive CNLOH (>20 arms), similar to previous reports for the rare thyroid Hürthle cell carcinoma (24, 25). Exclusion of the four tumors with ≥ 20 SCNAs each yielded 239 total SCNAs: 69 (28.9%) at the chromosome level, of which 48 were deletions, and 170 (71.1%) sub-chromosomal, of which 106 were deletions (Table S6).

In multivariable models, radiation dose was related to the number of sub-chromosomal SCNAs ($P=3.5 \times 10^{-5}$), particularly deletions ($P=7.0 \times 10^{-4}$) but not gains ($P=0.32$) or CNLOH ($P=0.52$); radiation dose also was not related to the number of chromosomal SCNAs ($P=0.20$) (Table S4). The most frequent recurrent event was loss of 22q (n=47/353, 13.3%) (Fig. S18–S19), but occurrence of 22q deletions was not associated with radiation dose ($P=0.37$) (Table S4).

Drivers of PTC

We identified at least one candidate driver for 433 of 440 (98.4%) tumors (Fig. S20, Table S7) (13), with the majority (n=401; 92.6%) having a single candidate driver, underscoring the parsimony of events driving PTC carcinogenesis. We designated 429 (97.5%) drivers for analysis (Fig. 1). Over half the designated driver events (n=253, 59.0%) were mutations (SSVs; Table S8), predominantly activating point mutations in genes previously implicated in PTC. The most commonly mutated gene was *BRAF* (n=194, 45.2%), where all the mutations either were canonical *BRAF*^{V600E} substitutions (n=190) or disrupted the V600 sequence context (n=4). *RAS* genes were the next most commonly mutated (n=44, 10.3%), specifically *NRAS* (n=20, 4.7%), *HRAS* (n=15, 3.5%), and *KRAS* (n=9, 2.1%). Additional mutation drivers were identified in *TSHR* (n=6, 1.4%), *DICER1* (n=3, 0.7%), *APC* (n=2, 0.5%), *TSC1/TSC2* (n=2, 0.5%), and *NFE2L2* (n=2, 0.5%). In TCGA, 9.4% of PTC harbored *TERT* promoter mutations, often in older individuals (8), but only one individual with a *TERT* promoter mutation was observed in our study (age at PTC diagnosis=40.7 years, designated driver=*BRAF*^{V600E}).

Fusion drivers accounted for 176 (41.0%) PTC cases (Table S9). The most frequently involved genes were *RET* (n=73, 17.0%) as well as other receptor tyrosine kinase (RTK) genes, specifically *NTRK3* (n=36, 8.4%), *NTRK1* (n=13, 3.0%), *ALK* (n=12, 2.8%), and *LTK* (n=3, 0.7%). Additional fusion drivers included *BRAF* (n=20, 4.7%) and *PPARG* (n=13, 3.0%), as well as SVs that resulted in overexpression of *IGF2* or *IGF2BP3* (n=6, 1.4%). Of the 23 chromoplexy events described above, 16 generated driver fusions. All

22q deletions co-occurred with known driver mutations, most frequently *RAS* mutations ($P_{\text{heterogeneity}}=2.8 \times 10^{-10}$; $n=22/38$, 56.4%) (Table S10).

In multivariable analyses, fusion drivers in PTC were more common in individuals exposed to higher radiation dose ($P=6.6 \times 10^{-8}$) and in those diagnosed at younger ages ($P=5.4 \times 10^{-9}$) relative to those with mutation drivers (Fig. 3B, Table S4). There was a suggestion of a heterogeneous effect of dose by specific gene fusion ($P_{\text{heterogeneity}}=0.020$), with higher doses on average for PTCs with *IGF2/IGF2BP3* or *BRAF* fusion drivers, whereas the dose distribution did not differ significantly among mutation drivers (*BRAF*, *RAS*, other; $P_{\text{heterogeneity}}=0.17$) (Fig. 4). We extended our observations by inclusion of 45 non-overlapping individuals with PTC (excluding the 20 individuals already in our analyses) drawn from a previous Chernobyl study with known drivers identified with RNAseq (7); over half had doses ≥ 500 mGy (mean age at PTC=24.2 years, mean age at exposure=7.2 years, mean dose=1050 mGy) (Table S1). That smaller sample set also suggested a radiation dose-related increase in fusion drivers ($P=0.069$), which was consistent with the results from our study ($P_{\text{heterogeneity}}=0.90$; pooled analysis of fusion vs. mutation driver, adjusting for age at PTC, sex, and study: $P=4.6 \times 10^{-9}$).

Gene expression and methylation patterns

We conducted several analyses to assess whether gene expression and methylation patterns were related to radiation dose. First, unsupervised clustering analyses restricted to PTC tumor tissue yielded 5 mRNA clusters, 5 miRNA clusters, and 3 methylation clusters (Fig. S21–S22) (13). None of these clusterings were associated with radiation dose ($P_{\text{mRNA}}=0.85$; $P_{\text{miRNA}}=0.38$; $P_{\text{methylation}}=0.10$), but each closely correlated with the driver gene pathway ($P_{\text{mRNA}}=1.6 \times 10^{-64}$; $P_{\text{miRNA}}=1.0 \times 10^{-9}$; $P_{\text{methylation}}=6.4 \times 10^{-43}$) (Fig. 5A, Table S4, Table S11, Fig. S23), supporting the overriding importance of the driver for RNA expression patterns (10). Second, we identified three transcriptional patterns important in PTC based on the TCGA analysis (10): the *BRAF*^{V600E}-*RAS* score (BRS), estimating the degree to which the mRNA, miRNA, and methylation profiles resemble either *BRAF*^{V600E} or *RAS*-mutated PTC; the thyroid differentiation score (TDS), based on expression of 16 thyroid metabolism and function genes; and the ERK-activity score of 52 expressed genes responsive to MEK inhibition (13). As expected, the mRNA, miRNA, and methylation BRS scores were highly correlated with one another ($r=0.78$ – 0.92) (Table S12). Consistent with TCGA (10), the three different mRNA-based scores also were significantly correlated, particularly the BRS with both the TDS ($r=0.69$) and the ERK score ($r=-0.66$). None of the scores were associated with radiation dose after correction for multiple testing ($P_{\text{mRNA-BRS}}=2.1 \times 10^{-3}$; $P_{\text{miRNA-BRS}}=5.5 \times 10^{-3}$; $P_{\text{methylation-BRS}}=0.082$; $P_{\text{ERK}}=0.011$, $P_{\text{TDS}}=7.8 \times 10^{-3}$) (Table S4, Fig. S24A), whereas each was strongly related to driver gene pathway ($P < 1.0 \times 10^{-30}$ for all scores) (Fig. S24B).

To confirm the lack of association between radiation exposure and gene expression patterns, we conducted exploratory analyses of the differential expression of specific genes and gene sets by dose (13). In multivariable linear regression models adjusted for age at PTC, sex, and batch, the P_{adjusted} for dose was <0.05 for five genes (Fig. 5B, Table S13), with the smallest P-value ($P_{\text{adjusted}}=8.0 \times 10^{-3}$; \log_{10} -fold expression change/100 mGy=0.059)

for transfer RNA asparagine (anticodon GUU) (*TRNAN-GUU-2*), which is a target of the transcription factor *MYBL2*, a key regulator of cell cycle progression and apoptosis (26, 27). However, each of these associations were attenuated when the model was further adjusted for the driver gene pathway (Table S13). In contrast, over half the genes were differentially expressed ($P_{\text{adjusted}} < 0.05$) among the different driver gene pathways (Table S13). Despite previous reports suggesting radiation dose could be linked to *CLIP2* expression (28–30), no such relationship was observed in our substantively larger study, which included 33 overlapping samples from the previously-published analyses (29) (Fig. S25), either in the overall set of PTC cases ($P=0.42$, Fig. 5C) or in subsets defined by early age at radiation exposure (Fig. S26). An exploration of expression signatures through gene set enrichment analyses was pursued in the Molecular Signatures Database (MSigDB; <https://www.gsea-msigdb.org/gsea/msigdb>) (31). For 3,213 gene sets (those related to hallmark biological processes, thyroid, radiation, and the genes included in germline analyses below) (13), multivariable regression models adjusting for age at PTC, sex, and batch revealed similar results as those above for single gene differential expression analyses, namely no gene set expression patterns were significantly associated with radiation dose ($P_{\text{adjusted}} < 0.05$), whereas over half were strongly associated with the driver gene pathway (Fig. S27, Table S14).

Germline genetic variation

Possible contribution of germline genetic variation to radiation-related PTC was investigated in individuals of comparable Ukrainian ancestry ($n=383$ individuals, including 305 exposed, 78 unexposed) (Fig. S28). Twelve previously reported risk SNPs for sporadic PTC were used to generate a polygenic risk score (PRS) (32). Multivariable analyses adjusting for population substructure revealed that unexposed individuals with PTC and those who received lower radiation doses were more likely to have higher genetic risk ($P=4.7 \times 10^{-4}$) (Fig. 6, Table S4). Analyses of the 12 individual SNPs, albeit underpowered, yielded three possible associations with radiation dose: rs1588635 (9q22.33; $P=0.012$), rs2289261 (15q22.33, *SMAD3*; $P=0.030$) and rs10069690 (5p15.33, *TERT*; $P=0.054$) (Table S15).

Investigation of rare potentially protein-damaging variants in genes and pathways related to thyroid or other cancer predisposition, clinical radiation sensitivity syndromes, and DNA damage response revealed no major differences in the burden of these variants among individuals who developed PTC after different radiation doses (Table S4, Table S16–S17). Only four individuals ($n=2$, 2.6% unexposed; $n=2$, 1.2% <100 mGy; 0, 0% 100 mGy) carried potentially protein-damaging variants in known thyroid cancer susceptibility genes (Table S16–S17).

Detailed analyses of molecular characteristics associated with radiation dose

Analyses by clonality for each of the deletion metrics (total deletion count, deletion:SNV ratio, and ID5 and ID8 mutational patterns) revealed that the radiation dose-related associations were restricted consistently to clonal rather than subclonal deletions (Fig. 2, Table S18).

Similarly, analyses of SCNAs also demonstrated associations only with clonal but not subclonal sub-chromosomal deletions (Table S18). Because distinct repair mechanisms can generate deletions of different lengths (33–35), we further stratified the clonal deletion count by length (Fig. S29) and found the strongest association between radiation dose and 5 bp clonal deletions with patterns characteristic of end-joining repair ($P=4.9\times 10^{-31}$) (Fig. 3C, Table S18). These results are consistent with the ID8 mutational association and suggest a key role for end-joining mechanisms in repairing radiation-induced DNA double-strand breaks. Analyses of the 5 bp clonal deletions by the amount of microhomology at the deletion boundary revealed consistent associations between radiation dose and deletions with 0–1 bp microhomology as well as those with 2 bp microhomology (Table S18). These results implicate non-homologous end-joining (NHEJ) repair mechanisms, which are employed regardless of the amount of microhomology, whereas alternative end-joining (alt-EJ) repair mechanisms such as theta-mediated end-joining (TMEJ) typically generate deletions with 2 bp microhomology (33–35). In an ancillary analysis, we quantified in the small insertions the number of TINS (locally templated insertions), which are characteristic of TMEJ repair (34), and found that TINS were not associated with radiation dose ($P=0.69$) (Fig. S30). Further exploration of insertions and deletions by genomic sequence context (8) revealed only weak correlations for radiation dose with occurrence of deletions classified by flanking GC content ($P=0.015$), proximity to CPG islands ($P=0.014$), and the mean replication timing at the variant locus ($P=0.010$); in each case deletions in the higher radiation dose groups were more similar to a random background distribution (Table S19) (13). No such correlations between radiation dose and genomic sequence context were observed for insertions.

We undertook similar analyses of SVs after confirming each event (Table S6) (13), identifying those SVs with <20 bp of intervening loss/gain at the breakpoint, which indicates repair by end-joining mechanisms (33–35). Increased radiation dose was strongly associated with simple/balanced SVs that were clonal ($P=1.4\times 10^{-16}$) but not subclonal ($P=0.91$), with a pronounced association for clonal simple/balanced SVs enriched for patterns characteristic of end-joining mechanisms ($P=5.5\times 10^{-19}$) (Fig. 3D) versus other clonal SVs ($P=0.41$, Fig. 3E) (Table S18). Further analyses demonstrated consistent associations for radiation dose with clonal simple/balanced/end-joining SVs with <4 bp and 4–<20 bp of intervening loss/gain (Table S18). Similar to our observations for small deletions, these results specifically implicate the importance of NHEJ repair, which accounts for almost all <4 bp events but which could contribute regardless of the amount of intervening loss/gain. By comparison, alt-EJ repair mechanisms primarily give rise to events with 4 bp of intervening loss/gain (33–35). Additional analyses of clonal simple/balanced/end-joining SVs by type revealed a strong association between radiation dose and inversions ($P=3.6\times 10^{-14}$), consistent with a previous report (8), but also an association with translocations ($P=4.4\times 10^{-4}$) (Table S18).

For each of the radiation dose-associated variables, the results were similar when we restricted the study population to exposed individuals (Table S20). Albeit based on limited statistical power, further restriction to individuals with exposures 1–<500 mGy revealed consistent associations for dose only with the clonal deletion:SNV ratio, enrichment of fusion drivers, and presence of clonal simple/balanced/EJ SVs (Table S20). Linear-quadratic

and linear-exponential models of radiation dose generally did not improve the model fit compared with a linear model for any of the variables, except for clonal small deletions (total and restricted to 5 bp EJ deletions) (Table S20).

Notably, radiation dose-related increases in clonal deletions (particularly the deletion:SNV ratio, ID8, and the number of clonal 5 bp EJ small deletions), as well as fusion PTC drivers were substantially more pronounced for individuals exposed at younger ages (Fig. 7, Table S21–S22), albeit based on small numbers for certain analyses. In contrast, the radiation dose-related increase in SCNA clonal sub-chromosomal deletions was most pronounced at longer latencies (Table S21).

Radiation-related acceleration of PTC development

Exploratory analyses to address previous reports that ionizing radiation exposure accelerates aging and cancer development (36, 37) revealed no such evidence in our study population. First, we stratified analyses of the relationship of clock-like SBS and ID signatures with age at PTC and latency but found no effect modification by radiation dose (age at PTC: $P_{SBS}=0.63$, $P_{ID}=0.93$; latency: $P_{SBS}=0.28$, $P_{ID}=0.21$) (Fig. S31–S32). Additionally, radiation dose-dependent associations with key molecular characteristics did not appear to be strongly modified by latency, after accounting for age at exposure and age at PTC (Table S21). Analyses of relative telomere length demonstrated the expected association between decreased telomere length and increased age at PTC in blood ($P=3.2\times 10^{-5}$) but not in normal thyroid tissue ($P=0.99$) or PTC ($P=0.81$), and there was no association between relative telomere length and thyroid radiation dose ($P>0.4$ for all tissues). Methylation profiles were evaluated to estimate epigenetic age acceleration using two established metrics (38, 39). Regressing epigenetic age against chronological age in the non-tumor thyroid tissue and then comparing the residuals from this predicted age in the PTC tissue (13) revealed no association between age acceleration and radiation dose using either metric ($P>0.1$).

Discussion

Our large-scale integrated analyses of the genomic landscape of PTC that developed following the 1986 Chernobyl nuclear power plant accident provide consistent evidence that ionizing radiation-induced DNA damage, particularly double-strand breaks, represents an early carcinogenic event in thyroid tumorigenesis following radiation exposure. These findings substantially extend preliminary reports of radiation-related human tumor characteristics (8, 9) by integrating data from multiple platforms with large sample size and detailed radiation dose data. Increasing radiation dose was strongly associated with increased likelihood of fusion versus point mutation drivers, simple/balanced SVs, and small deletions, with the strongest associations observed for those that bore hallmarks of NHEJ repair and were clonal, particularly for individuals exposed at a young age. However, no unique radiation-related biomarker was identified. Together, our results indicate that thyroid tumorigenesis following radiation exposure results from DNA double-strand breaks in the genome that have an impact on key thyroid cell growth and differentiation genes, which in turn drive the expression and epigenetic characteristics of individual PTCs.

Most tumors had evidence for only a single, known oncogenic driver, which involved the MAPK pathway in nearly all cases, which is consistent with previously published studies of sporadic PTCs (10, 40, 41). These findings combined with the low mutational burden in thyroid tumors emphasize the efficiency of driver mutations in thyroid tumorigenesis even following ionizing radiation exposure, in contrast to other environmentally-driven cancers, such as cigarette smoking and lung adenocarcinoma or ultraviolet light and melanoma, that often require multiple drivers and have multiple subclones together with substantial somatic burden (23).

Based on multiple lines of evidence, our study demonstrates striking radiation dose-related increases in DNA double-strand breaks in human thyroid cancers developing after the Chernobyl accident, extending results from in vitro and animal radiobiological experiments (33–35). In contrast, the PTCs did not have evidence of radiation-related specific base mutations or clustered mutations (42). Cells with DNA double-strand breaks can recruit various repair mechanisms, each of which leaves characteristic evidence in the repaired sequence. A series of analyses consistently implicated NHEJ as the most important repair mechanism for the radiation dose-associated DNA double-strand breaks observed in the PTCs. While the importance of end-joining repair in human tumors has been reported previously (8), our detailed examination of the local sequence context for the SVs (including fusion drivers) and small deletions enabled us to identify that radiation dose was most clearly associated with NHEJ rather than alt-EJ or other repair mechanisms. The lack of association between radiation dose and TINS further demonstrated the lack of importance of alt-EJ mechanisms. The importance of NHEJ repair also was supported by the lack of significant association between radiation dose and mutation signatures associated with APOBEC, which preferentially targets intermediates in replication and repair by homologous recombination (43). Our results necessitate further research such as using genetically modified organoids (44, 45) to establish the causal role of radiation-related DNA double-strand breaks predominantly repaired by NHEJ in human carcinogenesis.

The role of radiation-related DNA damage as an early step in PTC carcinogenesis following the Chernobyl accident is further supported by the lack of association between radiation dose and PTC transcriptomic and epigenomic features, despite the use of various analytic approaches, including clustering, differential expression by gene or miRNA, and gene set enrichment analyses. With our large sample size, we did not confirm the previously reported association between radiation dose and *CLIP2* expression (28–30), even when we restricted our analyses to individuals exposed at younger ages. Notably, however, the PTC transcriptomic and epigenomic features differed strikingly by driver gene/pathway, supporting the importance of the specific driver in shaping the tumor profile (10, 40, 41). Utilization of both WGS and RNA-seq enabled us to identify a driver in 98% of the PTCs in our study. Deletion of chromosome 22q has been suggested as a driver for PTC, but all cases in our study with 22q deletions also had other known PTC drivers, suggesting that 22q did not act independently in our set of individuals who developed PTC during young adulthood. Intriguingly, however, 22q deletions were strongly related to the driver pathway, occurring most commonly in *RAS*-mutated PTCs, suggesting that 22q deletion could provide a growth advantage or otherwise enhance the effect of certain MAPK drivers.

With our large sample size, we were able to explore the independent effects of radiation dose, age at PTC, age at exposure, and latency on PTC molecular characteristics. The pronounced evidence of radiation-related damage that we observed for individuals exposed at younger ages is consistent with epidemiologic analyses that have identified higher thyroid cancer risks with radiation exposure at younger ages (17). The relationship of a number of molecular characteristics, particularly total mutational burden and driver type, with age at PTC warrants further investigation across a broader age range (10). Additional research with detailed dose data is needed to understand whether our findings extend across a broader dose range, to other types of radiation, as well as to other tumor types, and whether radiation-related genomic characteristics have an impact on histopathological parameters (46–48). It has been hypothesized that ionizing radiation exposure could accelerate tumor development, and substantial evidence demonstrates that cancer survivors exposed to high-dose radiotherapy exhibit an aging phenotype (36, 37). However, exploratory analyses within our data did not support this hypothesis.

Our results have important implications for radiation protection and public health, particularly for low dose exposure, from two perspectives. First, the lack of a unique radiation-related pattern of molecular characteristics in the PTCs in our study, due in part to the random nature of ionizing radiation-related damage across the genome as well as the fact that other mutagens can cause DNA double-strand breaks, suggests that we are yet to establish a reliable biomarker to distinguish tumors induced by radiation versus other causes. Nevertheless, further investigation is warranted to consider how the radiation dose-associated characteristics we identified could be incorporated into estimates of the probability that a specific papillary thyroid tumor was caused by ^{131}I exposure (probability of causation [POC] (Fig. S5) (49, 50)), which is currently based on prior epidemiologic studies (17). Second, our data are consistent with a linear dose-response for the key molecular characteristics associated with radiation dose in the range examined in our analysis (< 1 Gy), which aligns with the extensive radiobiological literature and other epidemiologic evidence regarding DNA damage and cancer risk following ionizing radiation exposure (51, 52).

Our study population included a substantial number of PTCs occurring after < 100 mGy exposure, likely reflecting both the availability of samples from the Chernobyl Tissue Bank as well as the increased detection of pre-existing PTC in the population that may not become clinically evident until later, if at all, due to intensive screening and heightened awareness of thyroid cancer risk in Ukraine. The increased genetic risk based on the PRS was notable among PTCs that occurred after lower doses despite limited statistical power to investigate germline genetic variants. The low overall mutational burden of early adulthood PTC, small sample sizes in certain population subgroups, and uncertainties in radiation dose estimates also limited our statistical power to thoroughly investigate the shape of the dose-response curve, precisely identify the magnitude of radiation-related effects (as reflected by the wide confidence intervals for many effect estimates), or reliably identify new radiation signatures.

In conclusion, we have characterized the genomic landscape of PTC, the most frequent cancer observed after the Chernobyl nuclear accident. Our results demonstrate a dose-dependent carcinogenic effect of radiation derived primarily from DNA double-strand

breaks repaired by NHEJ that initiate subsequent thyroid tumor growth, the patterns of which are shaped not by radiation exposure but rather by the specific driver gene. The consistency of the spectrum of PTC drivers in our study population compared with previous PTC series suggests that current therapeutic approaches for PTC are appropriate even for tumors that arise following radiation exposure (53). Our work provides a foundation for further investigation of radiation-induced cancer, particularly with respect to differences in risk as a function of both dose and age, and underscores the deleterious consequences of ionizing radiation exposure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the commitment of the staff of the Laboratory of Morphology of Endocrine System and the staff of the Department of Surgery of Endocrine System of IEM, who prepared the pathological material for the study and operated on the patients; the confirmation of thyroid tumor diagnoses provided by the International Pathology Panel of the Chernobyl Tissue Bank, including Professors A. Abrosimov, T. Bogdanova, G. Fadda, J. Hunt, M. Ito, V. Livolsi, J. Rosai, and E.D. Williams, and Dr. N. Dvinskyh; Nathan Appel (Information Management Services, Inc., Calverton, MD) for programming support; Elizabeth C. Sasse (Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute) for support in creating figures and editorial review of the manuscript; Dr. Clara Bodelon (Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute) for consultation regarding methylation age calculations; Dr. Jia Liu (Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research) for consultation regarding analyses of mRNA expression and assessment of cross-platform concordance; Dr. Bin Zhu (Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute) for consultation regarding analyses of mRNA expression; Cameron Palmer (Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research) for assistance computing principal components for ancestry analyses; Timothy Myers (Laboratory of Genetic Susceptibility, Division of Cancer Epidemiology and Genetics, National Cancer Institute) for independent review of SV sequence context; and the leadership of The Cancer Genome Atlas for conduct of an initial pilot study. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). The opinions expressed by the authors are their own, and this material should not be interpreted as representing the official viewpoint of the U.S. Department of Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

Funding:

This project was supported by the Intramural Program of the National Cancer Institute, National Institutes of Health. The Chernobyl Tissue Bank is supported by the National Cancer Institute (U24CA082102).

REFERENCES AND NOTES

1. United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), "Sources and Effects of Ionizing Radiation: UNSCEAR 2008 Report to the General Assembly, with Scientific Annexes," (United Nations Publication, New York, 2010).
2. Tronko M et al. , Thyroid neoplasia risk is increased nearly 30 years after the Chernobyl accident. *Int J Cancer* 141, 1585–1588 (2017). [PubMed: 28662277]
3. Furukawa K et al. , Long-term trend of thyroid cancer risk among Japanese atomic-bomb survivors: 60 years after exposure. *Int J Cancer* 132, 1222–1226 (2013). [PubMed: 22847218]
4. Veiga LH et al. , Thyroid cancer after childhood exposure to external radiation: an updated pooled analysis of 12 studies. *Radiat Res* 185, 473–484 (2016). [PubMed: 27128740]
5. International Agency for Research on Cancer (Lyon, France, 2012).

6. Goh VST et al. , Construction of fluorescence in situ hybridization (FISH) translocation dose-response calibration curve with multiple donor data sets using R, based on ISO 20046:2019 recommendations. *Int J Radiat Biol* 95, 1668–1684 (2019). [PubMed: 31550191]
7. Liu QJ et al. , Assessment of retrospective dose estimation, with fluorescence in situ hybridization (FISH), of six victims previously exposed to accidental ionizing radiation. *Mutat Res Genet Toxicol Environ Mutagen* 759, 1–8 (2014). [PubMed: 24246721]
8. Behjati S et al. , Mutational signatures of ionizing radiation in second malignancies. *Nat Commun* 7, 12605 (2016). [PubMed: 27615322]
9. Efanov AA et al. , Investigation of the relationship between radiation dose and gene mutations and fusions in post-Chernobyl thyroid cancer. *J Natl Cancer Inst* 110, 371–378 (2018). [PubMed: 29165687]
10. Cancer Genome Atlas Research Network, Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 159, 676–690 (2014). [PubMed: 25417114]
11. Thomas GA, The Chernobyl Tissue Bank: integrating research on radiation-induced thyroid cancer. *J Radiol Prot* 32, N77–80 (2012). [PubMed: 22394998]
12. Thomas GA, Williams ED, Thyroid tumor banks. *Science* 289, 2283 (2000). [PubMed: 11041794]
13. Materials and methods are available as supplementary materials at the Science website
14. Likhtarov I et al. , Thyroid cancer study among Ukrainian children exposed to radiation after the Chernobyl accident: improved estimates of the thyroid doses to the cohort members. *Health Phys* 106, 370–396 (2014). [PubMed: 25208014]
15. Likhtarov I et al. , Reconstruction of individual thyroid doses to the Ukrainian subjects enrolled in the Chernobyl Tissue Bank. *Radiat Prot Dosimetry* 156, 407–423 (2013). [PubMed: 23595409]
16. Likhtarov I et al. , Estimation of the thyroid doses for ukrainian children exposed in utero after the chernobyl accident. *Health Phys* 100, 583–593 (2011). [PubMed: 22004928]
17. Ron E et al. , Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res* 141, 259–277 (1995). [PubMed: 7871153]
18. Berrington de Gonzalez A et al. , Epidemiological studies of low-dose ionizing radiation and cancer: rationale and framework for the monograph and overview of eligible studies. *J Natl Cancer Inst Monogr* 2020, 97–113 (2020). [PubMed: 32657348]
19. Hauptmann M et al. , Epidemiological studies of low-dose ionizing radiation and cancer: summary bias assessment and meta-analysis. *J Natl Cancer Inst Monogr* 2020, 188–200 (2020). [PubMed: 32657347]
20. Alexandrov LB et al. , The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101 (2020). [PubMed: 32025018]
21. Maruvka YE et al. , Analysis of somatic microsatellite indels identifies driver events in human tumors. *Nat Biotechnol* 35, 951–959 (2017). [PubMed: 28892075]
22. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR, Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 3, 246–259 (2013). [PubMed: 23318258]
23. ICGC TCGA Pan-Cancer Analysis of Whole Genomes Consortium, Pan-cancer analysis of whole genomes. *Nature* 578, 82–93 (2020). [PubMed: 32025007]
24. Gopal RK et al. , Widespread chromosomal losses and mitochondrial DNA alterations as genetic drivers in Hurthle cell carcinoma. *Cancer Cell* 34, 242–255 e245 (2018). [PubMed: 30107175]
25. Ganly I et al. , Integrated genomic analysis of Hurthle cell cancer reveals oncogenic drivers, recurrent mitochondrial mutations, and unique chromosomal landscapes. *Cancer Cell* 34, 256–270 e255 (2018). [PubMed: 30107176]
26. Lowe TM, Eddy SR, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955–964 (1997). [PubMed: 9023104]
27. Lachmann A et al. , ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 26, 2438–2444 (2010). [PubMed: 20709693]
28. Hess J et al. , Gain of chromosome band 7q11 in papillary thyroid carcinomas of young patients is associated with exposure to low-dose irradiation. *Proc Natl Acad Sci U S A* 108, 9595–9600 (2011). [PubMed: 21606360]

29. Selmansberger M et al. , CLIP2 as radiation biomarker in papillary thyroid carcinoma. *Oncogene* 34, 3917–3925 (2015). [PubMed: 25284583]
30. Selmansberger M et al. , Dose-dependent expression of CLIP2 in post-Chernobyl papillary thyroid carcinomas. *Carcinogenesis* 36, 748–756 (2015). [PubMed: 25957251]
31. Subramanian A et al. , Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550 (2005). [PubMed: 16199517]
32. Gudmundsson J et al. , A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun* 8, 14517 (2017). [PubMed: 28195142]
33. Chang HHY, Pannunzio NR, Adachi N, Lieber MR, Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 18, 495–506 (2017). [PubMed: 28512351]
34. Carvajal-Garcia J et al. , Mechanistic basis for microhomology identification and genome scarring by polymerase theta. *Proc Natl Acad Sci U S A* 117, 8476–8485 (2020). [PubMed: 32234782]
35. Zhao B, Rothenberg E, Ramsden DA, Lieber MR, The molecular basis and disease relevance of non-homologous DNA end joining. *Nat Rev Mol Cell Biol* 21, 765–781 (2020). [PubMed: 33077885]
36. Sehl ME, Carroll JE, Horvath S, Bower JE, The acute effects of adjuvant radiation and chemotherapy on peripheral blood epigenetic age in early stage breast cancer patients. *NPJ Breast Cancer* 6, 23 (2020). [PubMed: 32566744]
37. Tong J, Hei TK, Aging and age-related health effects of ionizing radiation. *Radiat Med Protect* 1, 15–23 (2020).
38. Horvath S, DNA methylation age of human tissues and cell types. *Genome Biol* 14, R115 (2013). [PubMed: 24138928]
39. Hannum G et al. , Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49, 359–367 (2013). [PubMed: 23177740]
40. Tufano RP, Teixeira GV, Bishop J, Carson KA, Xing M, BRAF mutation in papillary thyroid cancer and its value in tailoring initial treatment: a systematic review and meta-analysis. *Medicine (Baltimore)* 91, 274–286 (2012). [PubMed: 22932786]
41. Liang J et al. , Genetic landscape of papillary thyroid carcinoma in the Chinese population. *J Pathol* 244, 215–226 (2018). [PubMed: 29144541]
42. Turner C et al. , Human genetic disease caused by de novo mitochondrial-nuclear DNA transfer. *Hum Genet* 112, 303–309 (2003). [PubMed: 12545275]
43. Roberts SA et al. , Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell* 46, 424–435 (2012). [PubMed: 22607975]
44. Drost J, Clevers H, Organoids in cancer research. *Nat Rev Cancer* 18, 407–418 (2018). [PubMed: 29692415]
45. Drost J et al. , Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* 358, 234–238 (2017). [PubMed: 28912133]
46. LiVolsi VA et al. , The Chernobyl thyroid cancer experience: pathology. *Clin Oncol (R Coll Radiol)* 23, 261–267 (2011). [PubMed: 21333507]
47. Bogdanova TI et al. , Papillary thyroid carcinoma in Ukraine after Chernobyl and in Japan after Fukushima: different histopathological scenarios. *Thyroid*, (2020).
48. Williams ED et al. , Thyroid carcinoma after Chernobyl latent period, morphology and aggressiveness. *Br J Cancer* 90, 2219–2224 (2004). [PubMed: 15150580]
49. Kocher DC et al. , Interactive RadioEpidemiological Program (IREP): a web-based tool for estimating probability of causation/assigned share of radiogenic cancers. *Health Phys* 95, 119–147 (2008). [PubMed: 18545036]
50. Interactive RadioEpidemiological Program Information - Version 5.7 Available at: <https://radiationcalculators.cancer.gov/irep/>. (Accessed November 26, 2019).
51. National Research Council, Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2 Washington, DC: The National Academies Press. 10.17226/11340. (2006).

52. Barnard S, Bouffler S, Rothkamm K, The shape of the radiation dose response for DNA double-strand break induction and repair. *Genome Integr* 4, 1 (2013). [PubMed: 23522792]
53. Schlumberger M, Leboulleux S, Current practice in patients with differentiated thyroid cancer. *Nat Rev Endocrinol* 17, 176–188 (2021). [PubMed: 33339988]
54. Stezhko VA et al. , A cohort study of thyroid cancer and other thyroid diseases after the chornobyl accident: Objectives, design and methods. *Radiation Research* 161, 481–492 (2004). [PubMed: 15038762]
55. Hatch M et al. , Thyroid cancer and benign nodules after exposure in utero to fallout from Chernobyl. *J Clin Endocrinol Metab* 104, 41–48 (2019). [PubMed: 30445441]
56. Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2, (2013).
57. Van der Auwera GA et al. , From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43, 11 10 11–11 10 33 (2013). [PubMed: 25431634]
58. Dai J, 10.5281/zenodo.4543812. (2021).
59. Dai J, 10.5281/zenodo.4543818. (2021).
60. Fortin JP et al. , Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* 15, 503 (2014). [PubMed: 25599564]
61. Cawthon RM, Telomere measurement by quantitative PCR. *Nucleic Acids Res* 30, e47 (2002). [PubMed: 12000852]
62. Gadalla SM et al. , Effect of recipient age and stem cell source on the association between donor telomere length and survival after allogeneic unrelated hematopoietic cell transplantation for severe aplastic anemia. *Biol Blood Marrow Transplant* 22, 2276–2282 (2016). [PubMed: 27641680]
63. Birger C et al. , FireCloud, a scalable cloud-based platform for collaborative genome analysis: Strategies for reducing and controlling costs. *bioRxiv*, 209494 (2017).
64. Cibulskis K et al. , Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31, 213–219 (2013). [PubMed: 23396013]
65. Benjamin D et al. , Calling Somatic SNVs and Indels with Mutect2. *bioRxiv*, 861054 (2019).
66. Saunders CT et al. , Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817 (2012). [PubMed: 22581179]
67. Kim S et al. , Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* 15, 591–594 (2018). [PubMed: 30013048]
68. Wala JA et al. , SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res* 28, 581–591 (2018). [PubMed: 29535149]
69. Cibulskis K et al. , ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* 27, 2601–2602 (2011). [PubMed: 21803805]
70. Danecek P et al. , The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158 (2011). [PubMed: 21653522]
71. Ramos AH et al. , Oncotator: cancer variant annotation tool. *Hum Mutat* 36, E2423–2429 (2015). [PubMed: 25703262]
72. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP, Variant Review with the Integrative Genomics Viewer. *Cancer Res* 77, e31–e34 (2017). [PubMed: 29092934]
73. Smit A, Hubley R, Green P (RepeatMasker Open-4.0. 2013–2015 <<http://www.repeatmasker.org>>).
74. Costello M et al. , Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* 41, e67 (2013). [PubMed: 23303777]
75. Landau DA et al. , Mutations driving CLL and their evolution in progression and relapse. *Nature* 526, 525–530 (2015). [PubMed: 26466571]
76. Ellrott K et al. , Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst* 6, 271–281 e277 (2018). [PubMed: 29596782]

77. Rheinbay E et al. , Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* 578, 102–111 (2020). [PubMed: 32025015]
78. Kent WJ, BLAT--the BLAST-like alignment tool. *Genome Res* 12, 656–664 (2002). [PubMed: 11932250]
79. Taylor-Weiner A et al. , DeTiN: overcoming tumor-in-normal contamination. *Nat Methods* 15, 531–534 (2018). [PubMed: 29941871]
80. Drier Y et al. , Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res* 23, 228–235 (2013). [PubMed: 23124520]
81. Li Y et al. , Patterns of somatic structural variation in human cancer genomes. *Nature* 578, 112–121 (2020). [PubMed: 32025012]
82. Chen X et al. , Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222 (2016). [PubMed: 26647377]
83. Stewart C, 10.5281/zenodo.4558189. (2021).
84. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, Pan-cancer analysis of whole genomes. *Nature* 578, 82–93 (2020). [PubMed: 32025007]
85. Carter SL et al. , Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30, 413–421 (2012). [PubMed: 22544022]
86. Hartley SW, 10.5281/zenodo.4535549. (2021).
87. Lawrence MS et al. , Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501 (2014). [PubMed: 24390350]
88. Mermel CH et al. , GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 12, R41 (2011). [PubMed: 21527027]
89. Dawson ET, 10.5281/zenodo.4539528. (2021).
90. Dawson ET, 10.5281/zenodo.4539568. (2021).
91. Dawson ET, 10.5281/zenodo.4539636. (2021).
92. Kim J et al. , Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet* 48, 600–606 (2016). [PubMed: 27111033]
93. Purdue MP et al. , Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet* 43, 60–65 (2011). [PubMed: 21131975]
94. Cann HM et al. , A human genome diversity cell line panel. *Science* 296, 261–262 (2002). [PubMed: 11954565]
95. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL, GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. *G3 (Bethesda)* 9, 2447–2461 (2019). [PubMed: 31151998]
96. Price AL et al. , Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38, 904–909 (2006). [PubMed: 16862161]
97. Morton LM et al. , Subsequent neoplasm risk associated with rare variants in DNA damage response and clinical radiation sensitivity syndrome genes in the Childhood Cancer Survivor Study. *JCO Precis Oncol* 4, (2020).
98. Lek M et al. , Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291 (2016). [PubMed: 27535533]
99. Genomes Project Consortium et al. , A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
100. Cingolani P et al. , A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92 (2012). [PubMed: 22728672]
101. Landrum MJ et al. , ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42, D980–985 (2014). [PubMed: 24234437]
102. Liyanarachchi S et al. , Assessing thyroid cancer risk using polygenic risk scores. *Proc Natl Acad Sci U S A* 117, 5997–6002 (2020). [PubMed: 32132206]

103. Iyer MK, Chinnaiyan AM, Maher CA, ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* 27, 2903–2904 (2011). [PubMed: 21840877]
104. Benelli M et al. , Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* 28, 3232–3239 (2012). [PubMed: 23093608]
105. Haas BJ et al. , STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv*, 120295 (2017).
106. Wang K et al. , MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* 38, e178 (2010). [PubMed: 20802226]
107. Kumar S, Vo AD, Qin F, Li H, Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci Rep* 6, 21597 (2016). [PubMed: 26862001]
108. Liu S et al. , Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res* 44, e47 (2016). [PubMed: 26582927]
109. Steinberg MK, 10.5281/zenodo.4544030. (2021).
110. Lai Z et al. , VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* 44, e108 (2016). [PubMed: 27060149]
111. Steinberg MK, 10.5281/zenodo.4544031. (2021).
112. Dai J, 10.5281/zenodo.4543808. (2021).
113. Dai J, 10.5281/zenodo.4543824. (2021).
114. Benjamini Y, Hochberg Y, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B* 57, 289–300 (1995).
115. Aryee MJ et al. , Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369 (2014). [PubMed: 24478339]
116. Tian Y et al. , ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 33, 3982–3984 (2017). [PubMed: 28961746]
117. Pratilas CA et al. , (V600E) BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. *Proc Natl Acad Sci U S A* 106, 4519–4524 (2009). [PubMed: 19251651]
118. Dai J, 10.5281/zenodo.4543809. (2021).
119. P. T. C. Group et al. , Genomic basis for RNA alterations in cancer. *Nature* 578, 129–136 (2020). [PubMed: 32025019]
120. Hanzelmann S, Castelo R, Guinney J, GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14, 7 (2013). [PubMed: 23323831]
121. Imielinski M, Guo G, Meyerson M, Insertions and deletions target lineage-defining genes in human cancers. *Cell* 168, 460–472 e414 (2017). [PubMed: 28089356]
122. Panebianco F et al. , THADA fusion is a mechanism of IGF2BP3 activation and IGF1R signaling in thyroid cancer. *Proc Natl Acad Sci U S A* 114, 2307–2312 (2017). [PubMed: 28193878]
123. UNSCEAR, “Effects of ionizing radiation: United Nations Scientific Committee on the Effects of Atomic Radiation - UNSCEAR 2006 Report, Volume 1 - Report to the General Assembly, with Scientific Annexes A and B,” (United Nations, New York, 2008).
124. Preston D, Lubin J, Pierce D, McConney M, Shilnikova N, Epicure Risk Regression and Person-Year Computation Software: Command Summary and User Guide Ottawa: Risk Sciences International; 2015.

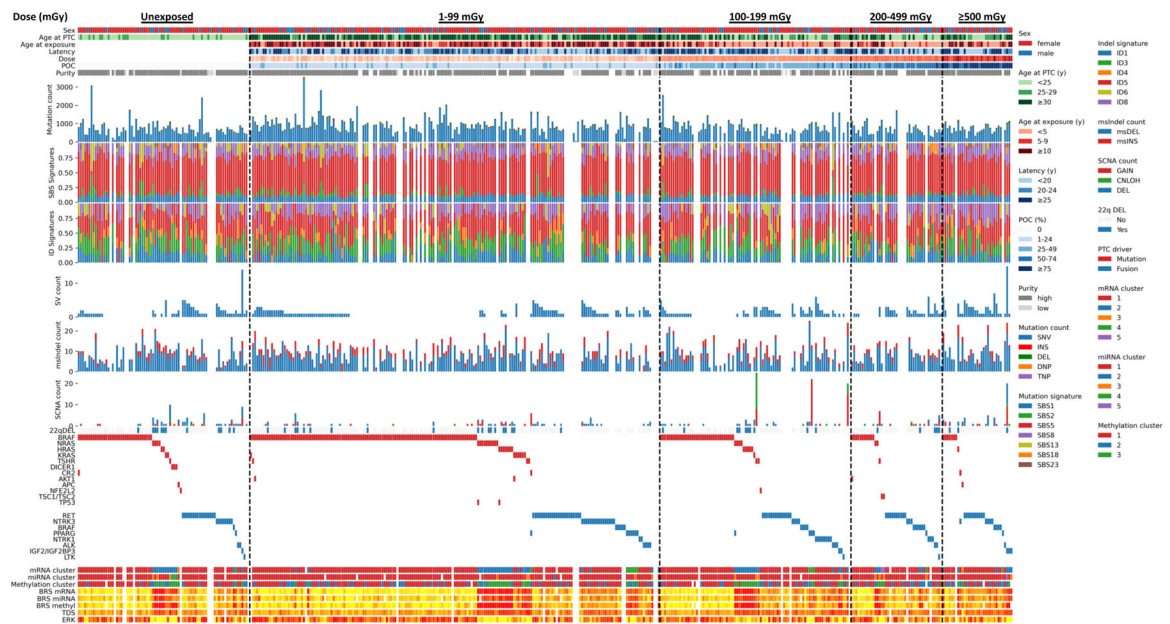


Fig. 1. Landscape of somatic alterations in 440 papillary thyroid carcinomas, by radiation dose from ^{131}I exposure.

Blank (white) spaces represent unavailable data due to lack of data from a specific platform (Figs. S1–S3). Signature analyses were restricted to high purity samples, defined as those with tumor purity $>20\%$ and no evidence of tumor contamination in the normal tissue. Abbreviations: *BRAF*^{V600E}-*RAS* score (BRS), copy neutral loss of heterozygosity (CNLOH), deletion (DEL), dinucleotide polymorphism (DNP, i.e., doublet), ERK-activity score (ERK), indel (ID), insertion (INS), microsatellite (ms), papillary thyroid carcinoma (PTC), probability of causation (POC), single nucleotide variant (SNV), somatic copy number alteration (SCNA), single nucleotide variant (SNV), structural variant (SV), thyroid differentiation score (TDS), trinucleotide polymorphism (TNP, i.e., triplet).

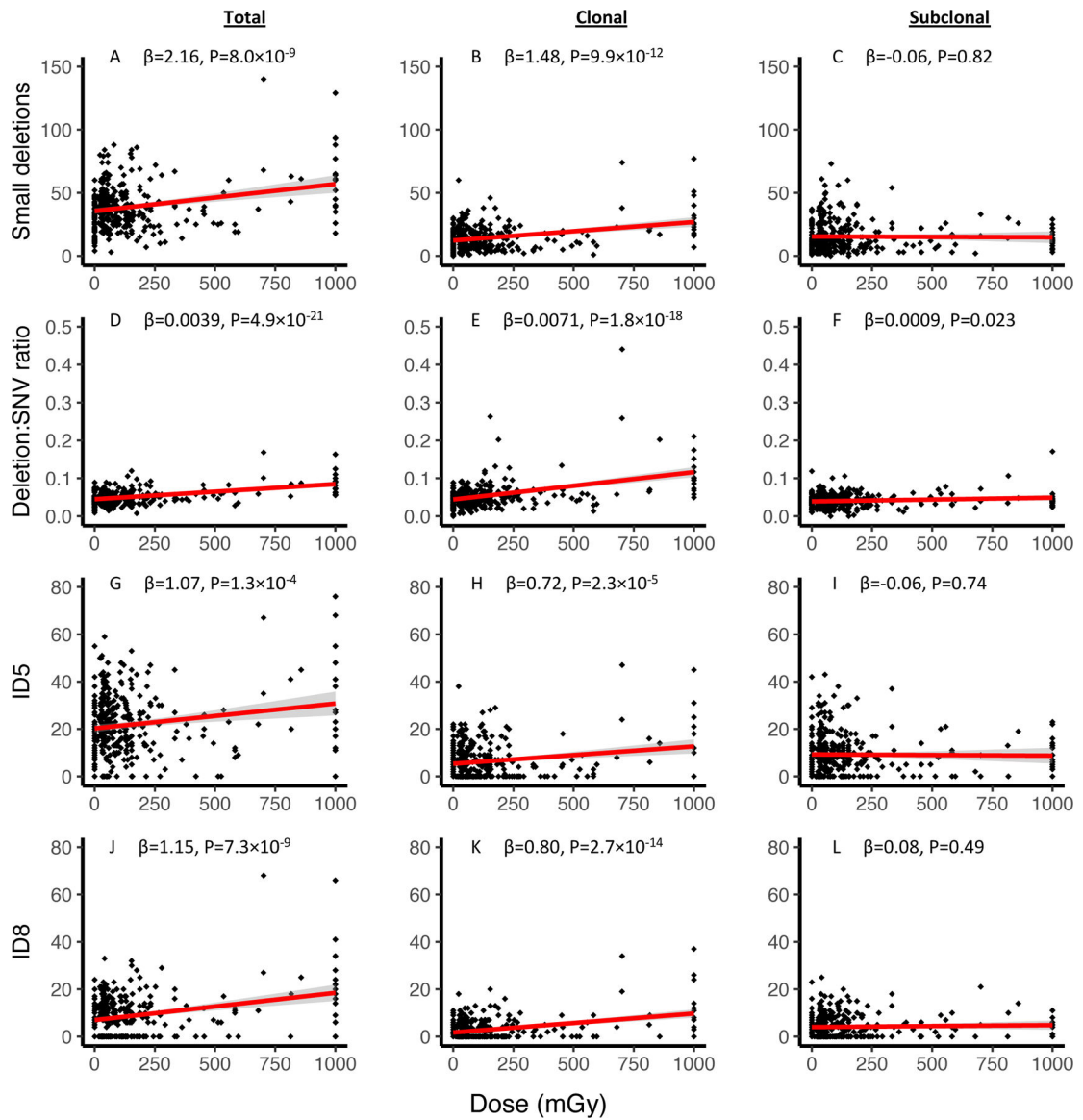


Fig. 2. Relationship between radiation dose from ^{131}I exposure and small deletions. (A) Total small deletion count and restricted to (B) clonal and (C) subclonal small deletions. (D) Total deletion:SNV ratio and restricted to (E) clonal and (F) subclonal deletions and SNVs. (G) Total ID5 count and restricted to (H) clonal and (I) subclonal ID5. (J) Total ID8 count and restricted to (K) clonal and (L) subclonal ID8. β per 100 mGy and P-value were derived from multivariable linear regression models adjusting for age at PTC and sex. Gray shading indicates 95% confidence interval (CI). Full model results are provided in Table S18.

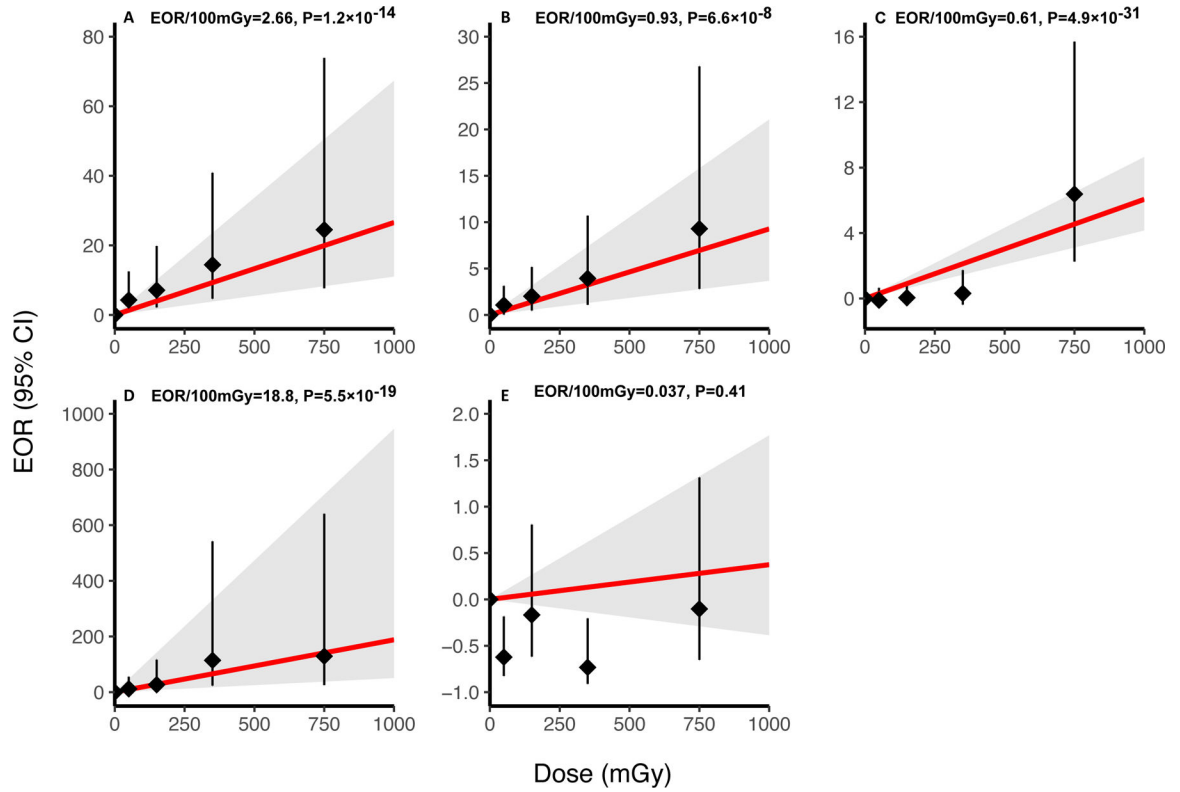


Fig. 3. Relationship between radiation dose from ¹³¹I exposure and selected SV metrics. (A) Number of simple/balanced SVs. (B) Likelihood of having a fusion versus mutation driver. (C) Number of clonal 5 bp EJ small deletions. (D) Number of confirmed clonal simple/balanced/end-joining SVs. (E) Number of confirmed clonal other SVs (E). Different scales are used for each panel to reflect the distributions and uncertainties of the excess odds ratio (EOR) estimates. Referent group for categorical analyses: EOR=0 (which is equivalent to odds ratio=1). EOR per 100 mGy and P-value were derived from multivariable proportional odds or logistic regression models adjusting for age at PTC and sex. Full model results are provided in Table S18.

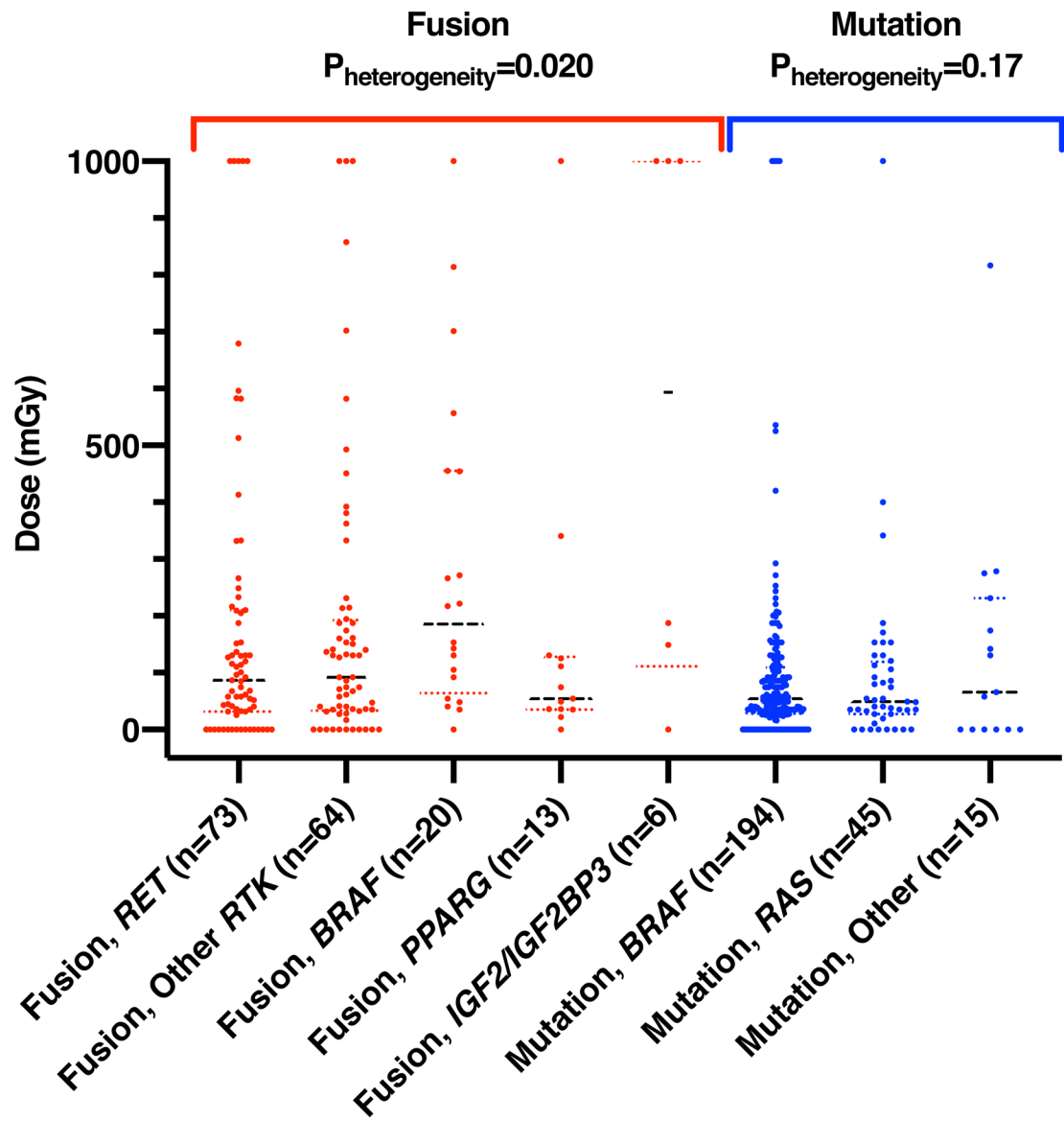


Fig. 4. Distribution of radiation dose from ^{131}I exposure by driver type and pathway.

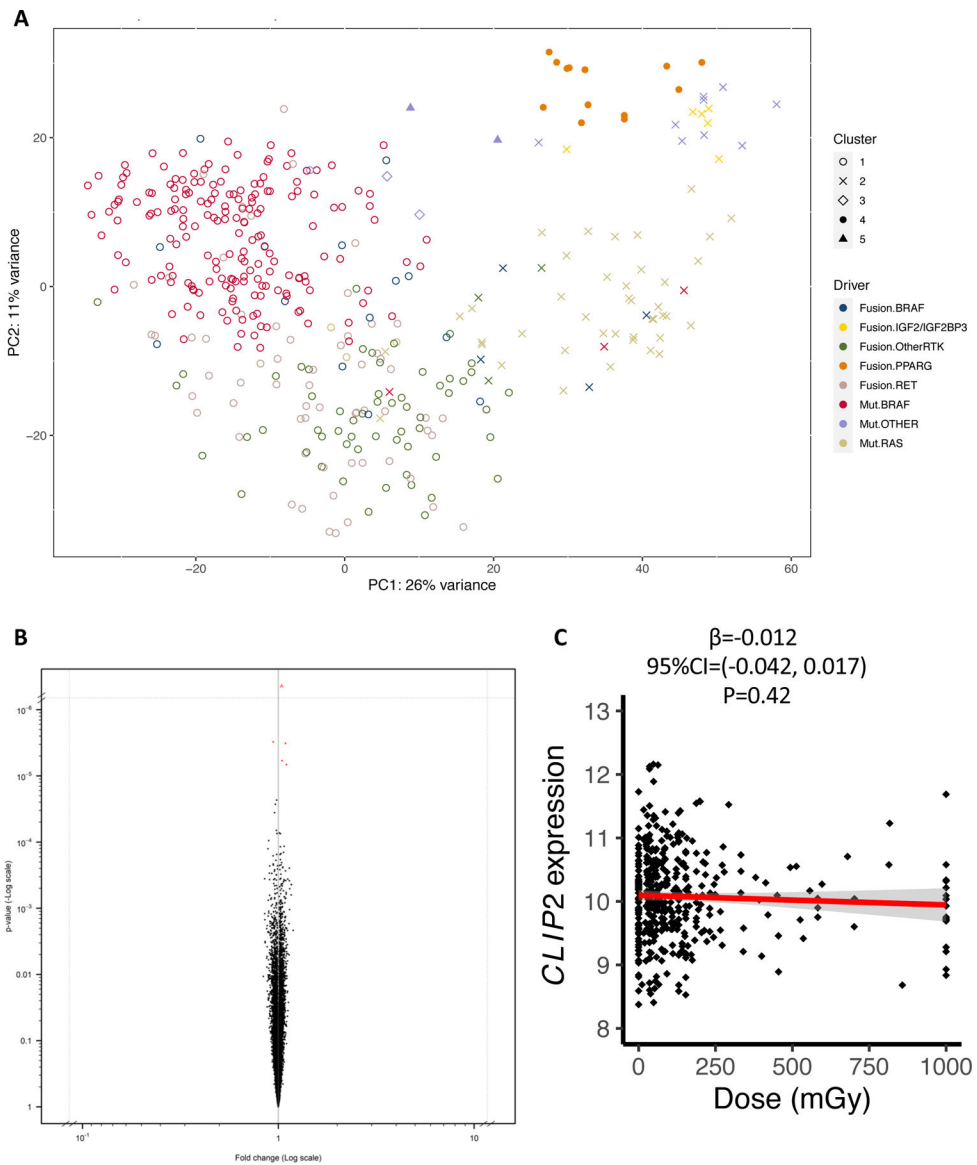


Fig. 5. Selected RNA-seq results.

(A) Differential expression by driver and cluster. (B) Differential expression for all genes by radiation dose from ¹³¹I exposure. (C) Differential expression of *CLIP2* by radiation dose from ¹³¹I exposure. Abbreviation: principal component (PC).

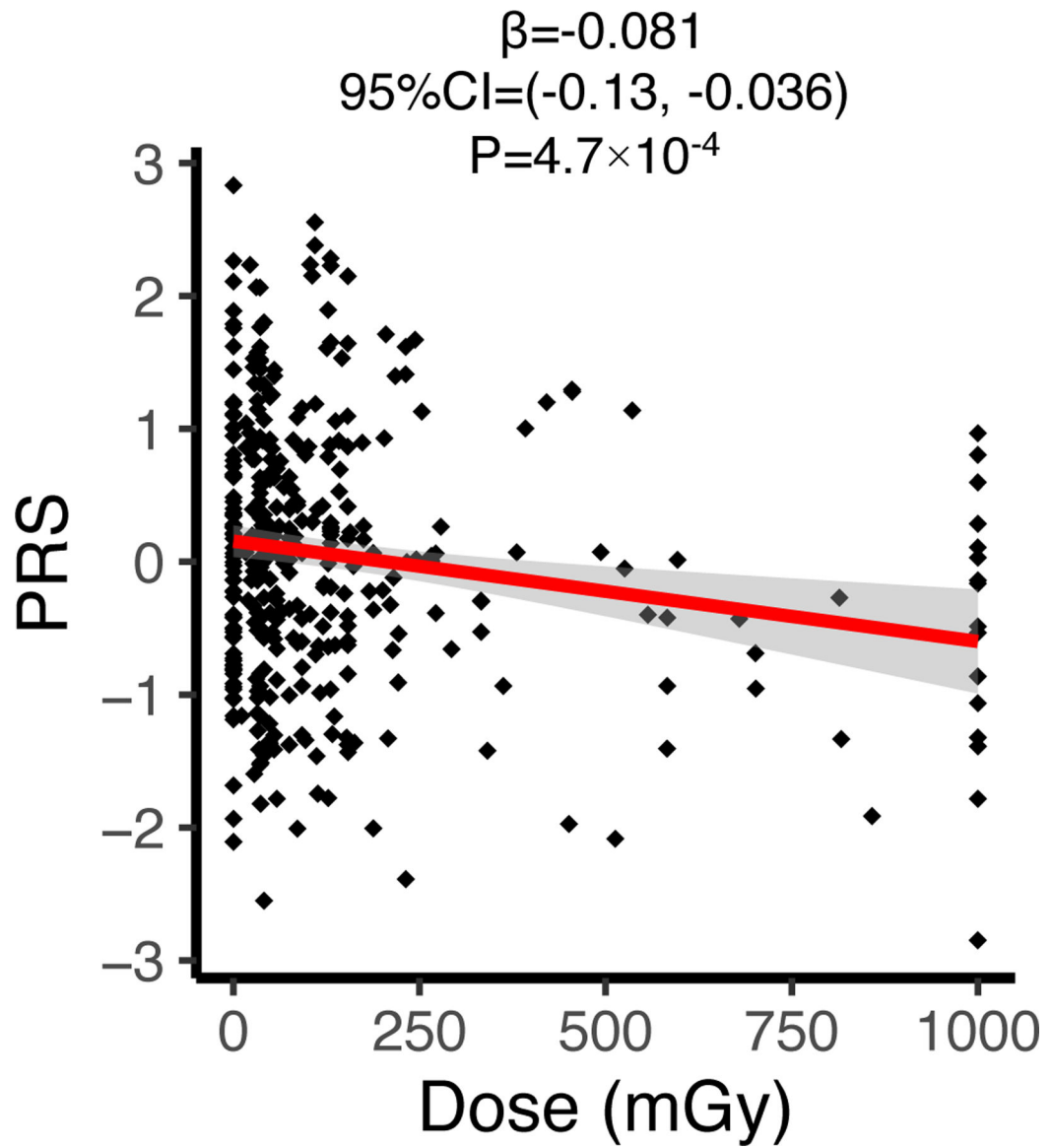


Fig. 6. Relationship between radiation dose from ^{131}I exposure and PRS.
Data for the 12 single nucleotide polymorphisms that comprise the PRS are provided in Table S15.

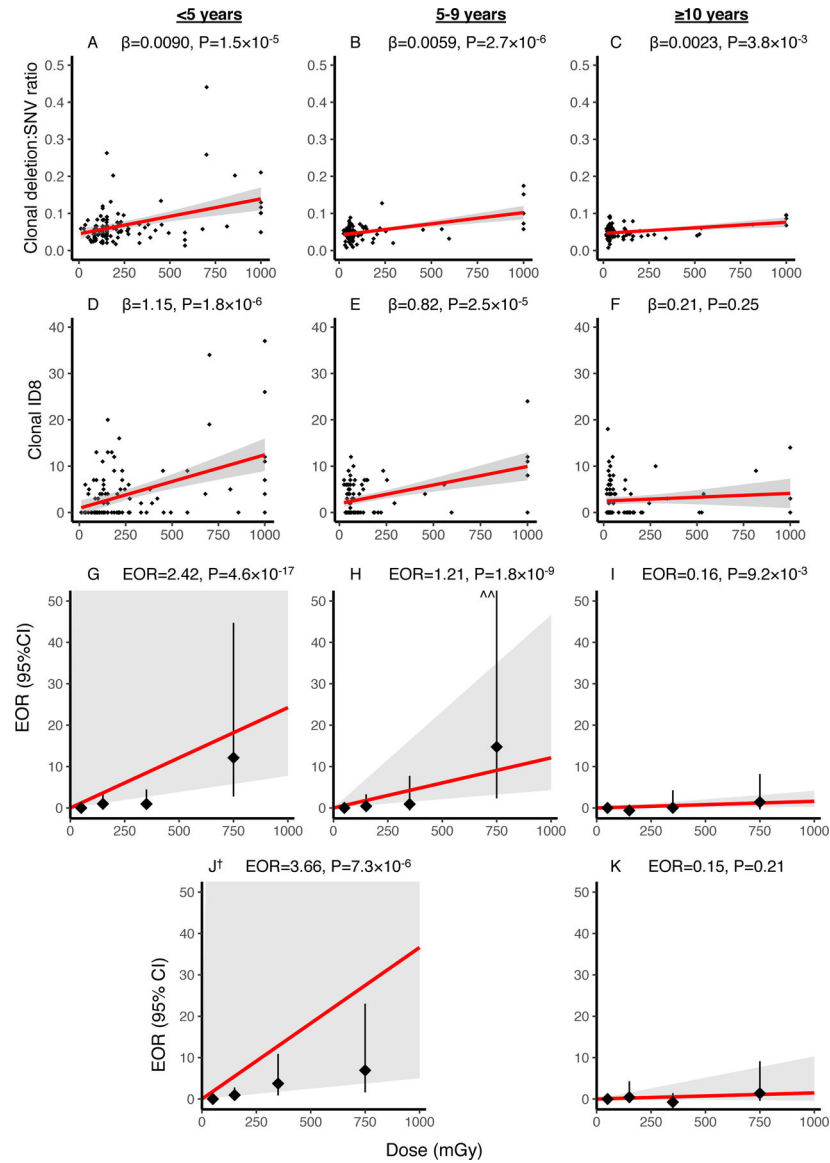


Fig. 7. Relationship between radiation dose from ^{131}I exposure and selected genomic characteristics by age at exposure. Clonal deletion:SNV ratio for (A) <5 years at exposure, (B) 5–9 years at exposure, and (C) 10 years at exposure. Number of clonal ID8 mutations for (D) <5 years at exposure, (E) 5–9 years at exposure, and (F) 10 years at exposure. Number of clonal 5 bp EJ small deletions for (G) <5 years at exposure, (H) 5–9 years at exposure, and (I) 10 years at exposure. Likelihood of having a fusion versus mutation driver for (J) <10 years at exposure and (K) 10 years at exposure. All analyses exclude ^{131}I -unexposed individuals. β or EOR per 100 mGy and P-value were derived from multivariable linear, proportional odds, or logistic regression models adjusting for age at PTC and sex. Full model results are provided in Table S22. * Models evaluating the effect of dose on driver type restricted to <5 years of age at exposure did not converge, so individuals exposed at <5 and 5–9 years

were combined in panel J. EOR/100 mGy for 5–9 years alone=1.78, 95%CI=0.12–226. ^^ in Panel H indicates that the upper 95% CI exceeds the y-axis maximum value.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript