

Integration of a radiation biomarker into modelling of thyroid carcinogenesis and post-Chernobyl risk assessment

Jan Christian Kaiser^{1,*}, Reinhard Meckbach², Markus Eidemüller¹, Martin Selmansberger³, Kristian Unger³, Viktor Shpak⁴, Maria Blettner⁵, Horst Zitzelsberger³, Peter Jacob⁶

¹Institute of Radiation Protection, Helmholtz Zentrum München, 85764 Oberschleißheim, Germany

²Boris-Blacher-Str. 14, 80939 München, Germany

³Helmholtz Zentrum München, Research Unit Radiation Cytogenetics, 85764 Neuherberg, Germany;

⁴National Academy of Medical Sciences of the Ukraine, Institute of Endocrinology and Metabolism, 254114 Kyiv, Ukraine;

⁵Johannes Gutenberg Universität, Institut für Medizinische Biometrie Epidemiologie und Informatik, 55131 Mainz, Germany

⁶RADRISK, 83727 Schliersee, Germany

*To whom correspondence should be addressed. Tel: 00498931874028; Fax: 004931873363

Email: christian.kaiser@helmholtz-muenchen.de

Abstract

Strong evidence for the statistical association between radiation exposure and disease has been produced for thyroid cancer by epidemiological studies after the Chernobyl accident. However limitations of the epidemiological approach in order to explore health risks especially at low doses of radiation appear obvious. Statistical fluctuations due to small case numbers dominate the uncertainty of risk estimates. Molecular radiation markers have been searched extensively to separate radiation- induced cancer cases from sporadic cases. The overexpression of the CLIP2 gene is the most promising of these markers. It was found in the majority of papillary thyroid cancers (PTCs) from young patients included in the Chernobyl tissue bank. Motivated by the CLIP2 findings we propose a mechanistic model which describes PTC development as a sequence of rate-limiting events in two distinct paths of CLIP2-associated and multi-stage carcinogenesis. It integrates molecular measurements of the dichotomous CLIP2 marker from 141 patients into the epidemiological risk analysis for about 13,000 subjects from the Ukrainian-American cohort which were exposed below age 19 yr and were put under enhanced medical surveillance since 1998. For the first time a radiation risk has been estimated solely from marker measurements. Cross checking with epidemiological estimates and model validation suggests that CLIP2 is a marker of high precision. CLIP2 leaves an imprint in the epidemiological incidence data which is typical for a driver gene. With the mechanistic model we explore the impact of radiation on the molecular landscape of PTC. The model constitutes a unique interface between molecular biology and radiation epidemiology.

Introduction

Radio-epidemiological studies provide evidence that exposure to ionizing radiation is especially harmful during childhood. Application for diagnostic or therapeutic purposes can cause diseases such as leukemia or solid cancers of the brain, breast or thyroid even decades after exposure (1-3).

Strongest evidence for the statistical association between radiation exposure and thyroid cancer has been produced by studies of children and adolescents who ingested radio-iodine from fallout of the Chernobyl accident in April 1986. Early studies were published about a decade after the event and have already reported a marked increase of the radiation risk (reviewed in (4,5)). In later studies on about 13,000 subjects of the Ukrainian-American (UkrAm) cohort, which were exposed below age 19 yr and were put under enhanced medical surveillance since 1998, the excess relative risk (ERR) per thyroid dose dropped from about 5 per Gy for mean age at operation (AaO) 16 yr to just below 2 per Gy for mean AaO 24 yr (6,7). The estimate for a similar cohort from Belarus is compatible with this decreasing trend of attained age (8). For exposure in adulthood the risk in Japanese a-bomb survivors is just marginally elevated (9).

Although conventional radio-epidemiological analysis is the method of choice to determine the dose-response relationship for a given disease, limitations of this approach become increasingly obvious. Statistical power can be improved by pooling of cohorts but uncertainties on dose estimates and other co-factors are accumulated as well (10). The accuracy of risk estimates even from studies with thousands of cases might not be sufficient to decide if radiation is causing detrimental health effects at doses below about hundred mGy (11).

Molecular biomarkers of ionizing radiation are being searched intensively to separate radiation-induced cancer cases from sporadic cases with greater accuracy than can be gained from epidemiology (12). But to date many such attempts lack a satisfactory justification because mechanisms that relate radiation to carcinogenic processes on a molecular level are still not fully explained (13). Based on integrative systems biology the search for a molecular biomarker of a disease caused by radiation is probably most advanced in thyroid carcinogenesis. The histological composition of more than 90% of all thyroid carcinoma from the Chernobyl studies is of the papillary type (PTC). Genomic analysis of PTC samples stored in the Chernobyl tissue bank (CTB, www.chernobyltissuebank.com) revealed a DNA copy number gain on chromosome band 7q11.23

and mRNA overexpression of the CLIP2 gene related to radiation exposure (14). The samples came from the Genrisk-T cohort of patients which were exposed at very young age below five years.

Radiation-related CLIP2 overexpression has been validated with independent samples from cohorts Genrisk-T and UkrAm at protein level using a dedicated protocol which combined immunohistochemistry with digital image analysis (15). Results from both studies suggest that CLIP2 overexpression can be applied as a dichotomous biomarker to decide on the sporadic or radiation-associated origin of a PTC in young patients operated below age 20 yr (16).

The strategic aim of the EU-funded EpiRadBio project was to improve risk assessment at low doses by benefiting from molecular measurements in cancer tissue from patients who were also members of radio-epidemiological cohorts. This aim is addressed here with a multi-scale approach of modelling which comprises data over several length orders from point mutations to cells and tumors (17). It enables us to relate the CLIP2 findings to the radiation risk of the UkrAm cohort and to explore the impact of radiation on molecular alterations such as mutations, genomic copy number alterations (CNAs) or altered mRNA and protein expression (18).

To harness molecular data as hallmarks of key carcinogenic processes for risk assessment we develop mechanistic models. Simple versions of such models have been developed in the 1980s and have later been couched in a general mathematical framework (19,20). They already treated carcinogenesis as a progression of cell-based events at yearly rates such as mutations on the pathway to cancer. But the success of mechanistic models remained limited since the association of model parameters with well-defined molecular changes appeared rather vague. Growing insight into molecular processes has been gained from numerous omics technologies. It now offers a more tangible characterization of the stages which are featured in a mechanistic model. Beneath mutations in driver genes they comprise a wider range of phenomena such as gene rearrangements or fusions and even chromosomal instabilities (gains and losses of chromosomes or chromosome arms) (21,22). Clonal growth of pre-cancerous lesions from cells which have accumulated initial genetic damage could also influence the incidence data. As general condition, molecular events must be rate-limiting to be detectable in cancer incidence data. As example, the model of radiation-induced colon cancer in Japanese a-bomb survivors includes already a number of such rate-limiting events (23).

In this work we investigate whether CLIP2 overexpression constitutes a reliable radiation marker which can improve risk assessment. Using mechanistic modelling as a tool, we present a characterization of clinical and molecular properties associated with sporadic and radiation-induced PTC.

Accepted Manuscript

Materials and methods

Data sets

UkrAm cohort for radiation epidemiology

Scientific aims, design and risk assessment for the UkrAm cohort are already presented in previous studies (6,7). The cohort data was gathered in the Ukraine during four screening cycles between 1998 and 2008. Previous analyses were performed with either 45 prevalent cases of the first screening cycle (6) or 65 incidental cases of the second to fourth screening cycles (7). In those studies all histological types of papillary, follicular and medullary thyroid cancer were combined for the main risk analysis. The present study is only concerned with development of PTC. To achieve maximal statistical power all 115 PTC cases were included which have been operated in the UkrAm cohort until 2008. The buildup of the cohort data for the present study is given in the Supplementary Material. A breakdown of thyroid cancer cases with their use in previous studies compared to the present study is shown in Table S1. The data set of 13,183 subjects for present study is summarized in Table S2. We applied the same individual dose estimates from dosimetry system TDose10 (24-26) as Brenner et al. (7). Model development and risk analysis has been performed with fits to the UkrAm cohort.

Tumor cohort for molecular biology

In contrast to the above-mentioned epidemiological UkrAm cohort, our tumor cohort of 141 PTC patients from the CTB with individual CLIP2 status and TDose10 estimates (27) has been formed previously by Selmansberger et al. (16). They have joined the discovery cohort Genrisk-T with validation cohorts Genrisk-T plus and UkrAm (CTB tumors only) which were investigated in earlier exploratory studies related to the CLIP2 marker (14,15). All PTC cases are listed online in their supplementary Table 1. 70 of our 141 cases from the CTB also belonged to the epidemiological UkrAm cohort (Table 1). Most of the young patients with AaO < 20 yr (63 out of 76) pertained to

cohorts Genrisk-T and Genrisk-T plus. Both cohorts also contained 24 patients with AaO < 20 yr and nominal doses of 1 mGy/yr who were born after 1986. Parameters of the mechanistic model were determined by fits to the epidemiological UkrAm cohort. But model results on the prediction of the CLIP status have been validated with our tumor cohort. For subsets of the tumor cohort measurements on various molecular alterations have been performed (18). The impact of radiation on these alterations has been investigated with the mechanistic model.

Mechanistic model of PTC pathogenesis

Biological basics of thyroid carcinogenesis

The human thyroid contains about 10^{10} follicular cells (thyrocytes) (28). Among them about 0.1% somatic stem cells might be susceptible to cancer development (29). Molecular and functional properties of somatic stem cells are largely unknown because they are rare and difficult to identify (30). The adult thyroid gland is a slowly proliferating organ with a turnover rate of 8.5 to 14.4 yr in order to maintain its size and function (31). PTC is the most common histological type of thyroid cancer. According to the paradigm of multi-stage carcinogenesis (MSC) PTCs develop from unknown precursor cells deriving from normal thyrocytes by acquiring genomic instability (32).

Rearrangements of the RET gene (RET/PTC) are frequent alterations in PTCs (33,34). However, their association with radiation exposure is not well established (35). Point mutations of the BRAF gene and RAS mutations are both mutually exclusive but occur mainly in sporadic PTC (18,36). Both mutations lead to a deregulated MAPK pathway. Furthermore, CNAs are frequently reported in PTCs and might be considered as hallmarks of MSC (18,32).

The oncogenic processes leading to PTC development are not fully understood. We assume that CLIP2 overexpression is an early event predominantly in radiation-related PTC development. A functional network analysis of the CLIP2 gene revealed associations with genes involved in both MAPK signaling and genomic instability (15). By triggering RET/PTC rearrangements CLIP2

overexpression may represent a driver event but a mere passenger event cannot be excluded. Compared to MSC we expect to find a lower number of CNAs in CLIP-related carcinogenesis (CSC) which also involves fewer genes (14,36-38). During a comparatively short latency period of less than 17 yr from exposure at very young age (< 5 yr) to surgery precancerous cells do not develop much differentiation.

Takano (39) has argued against the role of MSC in PTC development after Chernobyl because thyrocytes rarely proliferate. Neither are intermediate cells known, nor have morphological precursor lesions been detected. PTC appeared most frequently in patients exposed at very young age assuming that thyroblasts are still present as potential radiation targets. The disappearance of thyroblasts in older subjects therefore leads to a reduced radiation sensitivity of the thyroid gland. But the form of an age-dependent decay function for thyroblasts is not known. Takano's (39) notion of pathogenesis for post-Chernobyl PTC is represented by the molecular changes related to C2C.

Conceptual model

We can now construct a mechanistic model of carcinogenesis which reconciles the determinant biological features of both MSC and C2C. Different molecular mechanisms in early PTC development imply a conceptual model design which relies on distinct pathways for the pathogenesis of sporadic PTC without the CLIP2 biomarker (MSC) compared to PTC with the CLIP2 biomarker (C2C). As a common feature both pathways include deregulated MAPK signaling albeit at different stages (Figure 1). Our preferred mechanistic model consists of two submodels for MSC and C2C with separate mathematical formulations which are outlined in the Supplementary Material (Eqs. S1-S12).

The MSC submodel for sporadic PTC without the CLIP2 marker starts with healthy thyrocytes. It involves two initiating mutations with rates v_0 , v_1 followed by a phase of clonal expansion with net rate $\gamma = \alpha - \beta$. It is inspired by similar models for cancers in the colon and thyroid (23,40,41). The rates

for symmetrical stem cell division α and inactivation (or differentiation) β of initiated cells cannot be identified from the incidence data but the net rate γ of clonal expansion can be estimated quite well. Formally, the process is concluded when at least one cell from the premalignant clones suffers a late transforming event which produces a non-extinct tumor stem cell (40). This event with effective rate v_2 summarizes a series of complex processes in the late carcinogenic stage. The model for the MSC path reflects the pertinent hallmarks of cancer such as copy number alterations, mutations in driver genes and clonal expansion of precancerous lesions (21,22). However, statistical support for the latter process appears rather weak in the model.

The C2C submodel for young patients (AaO < 20 yr) is based on a hypothesis by Takano (39) who proposes a dedicated molecular path starting from fetal thyroid cells (or thyroblasts). A numerical value for the number N_b of thyroblasts is difficult to estimate. In young patients the CLIP2 marker is induced by radiation, whereas in a small number of older patients the CLIP2 marker appears also sporadically (16). We assume that C2C occurs in a simple two-stage model with different parameter dependencies for radiation-induced C2C (rC2C) and sporadic C2C (sC2C). In the C2C submodel two transitions (or mutational hits) with rates μ_0 , μ_1 are required to produce a cancer cell. Radiation action increases the first hit μ_0 for a short period to induce a permanently enhanced radiation risk after exposure. Sporadic PTCs within C2C are induced by a step-like increase of the second hit μ_1 after age 20 yr possibly due to altered cell interactions in the adult thyroid. This step is immediately reflected in the hazard for sC2C without delay.

Statistical analysis

To put goodness-of-fit and risk estimates of mechanistic models into perspective, descriptive risk models, which are based solely on statistical associations between exposure and disease, have also been applied to the epidemiological UkrAm cohort. The epidemiological analysis was performed with the MECAN software package where the risk models have been implemented (42). Details on the form of the likelihood function, parameter estimation and uncertainty analysis are given in the

Supplementary Material. Statistical analysis of the tumor cohort has been performed with the R package (43).

Accepted Manuscript

Results

Preferred risk models

The preferred descriptive and mechanistic risk models have been selected from about forty candidate models according to criteria of goodness-of-fit and biological plausibility. In Table S4 deviance and number of adjusted model parameters are shown for a shortlist of candidate models. The preferred models accidentally produced the same deviance. To correct for unexplained variations in PTC incidence possibly related to iodine deficiency the total risk for Zhytomir oblast has been adjusted for by a heuristic factor $\exp(\Phi_{\text{obl86}})$. Model parameters are listed in Tables S5 and S6.

Descriptive models of the excess absolute risk (EAR) yielded better fits than ERR models for the same adjustment for oblast. Thus, descriptive ERR estimates were calculated from the preferred EAR model described in Eq. (S13).

For the preferred mechanistic model the five adjustable parameters of the MSC and rC2C components were estimated by fits to the UkrAm cohort data (Table S5). The three parameters of the sC2C model component were determined separately after the parameters for the rC2C and MSC components have been fitted by likelihood regression. They were adjusted to produce the expected 19 PTCs with the CLIP2 marker in 86 patients with AaO ≥ 20 yr in the UkrAm cohort. The decomposition of the hazard function into the three components for MSC, rC2C and sC2C is shown in Figure S3.

Validation with the CLIP2 marker

A pertinent link between molecular biology and radiation epidemiology is established by comparing the conditional probability $P_{\text{mol}}(\text{C2C}|\text{PTC})$ of finding the CLIP2 marker in a PTC from molecular measurements with the epidemiological estimate $P_{\text{epi}}(\text{C2C}|\text{PTC})$ of the same quantity. The construction of P_{epi} from hazard functions of the preferred mechanistic model is given in the

Supplementary Material by Eq. (S19). In radiation epidemiology P_{epi} is often termed probability of causation.

In Figure 2 the estimate for P_{epi} at attained age 17 yr is compared with the response P_{mol} for AaO < 20 yr of Selmansberger et al. (16). For young patients the probabilities exhibit similar dose responses in view of their uncertainties. Older patients do not possess a dose response so that in Figure S4 only the crude rate of the positive CLIP2 marker among PTCs with AaO \geq 20 yr is indicated. For AaO \geq 20 yr P_{epi} shows an offset at zero dose which is caused by sporadic PTCs in the sC2C path.

To substantiate the visual impression of Figures 2 and S4 a comparison of measurements and model predictions on the CLIP status has been performed. The 141 PTCs from our tumor cohort were subdivided in three different dose groups for AaO < 20 yr and a single group for all PTCs with AaO \geq 20 yr combined (Table S7). The predicted numbers of CLIP-positive cases were calculated with Eq. (S21). The 95% CI of the measured crude rate included the rate from model prediction in all four groups.

We present the same comparison for a subdivision of our 141 cases in the discovery cohort Genrisk-T and in validation cohorts Genrisk-T plus and UkrAm (CTB tumors only) (15). In total the mechanistic model predicts 92/49 cases with CLIP2-positive/negative biomarker in compliance with measurements of 93/48 cases (16). Table 2 shows remarkable agreement between model predictions and measurements for AaO < 20 yr. Model predictions suggest for patients born after 1986 that even very low doses of background radiation cause PTC. For AaO \geq 20 yr the effect of the sC2C path appears which generated about 14 sporadic cases in addition to the 25 radiation-induced cases from the rC2C path.

Risk assessment

Risk estimates of the present study are derived solely for PTC whereas other risk studies report mostly estimates for all histological types of thyroid cancer. This difference is negligible since PTC is

the dominant type. Table S8 presents estimates of the EAR and the ERR from the preferred mechanistic and descriptive models. They are almost identical for both models so that a separate discussion is not necessary.

Excess absolute risk

We found no dependence on attained age, age at exposure or time since exposure for estimates of the EAR in contrast to estimates for a-bomb survivors which show a marked increase for young age at exposure (9,11,44). Up to 1 Gy the dose response is nearly linear, for higher doses exponential attenuation sets in. The oblast-averaged point estimate for both sexes of 5.1 PTC cases per 10,000 PY at 1 Gy in Table S8 is a factor of 2-3 higher compared to estimates from studies for Chernobyl children and a-bomb survivors (7,9,44,45). The reason for different estimates of the EAR from the present study and of 1.7 (95% CI < 0.06; 5.0) PTC cases per 10,000 PY at 1 Gy from Brenner et al. (7) is mainly related to their approach of oblast correction. The present approach lowers the deviance by about ten points compared to the approach of ref. (7). Our EAR estimate agrees well with the estimate of 4.4 (95% CI 1.9; 10.1) cases per 10,000 PY at 1 Gy from Ron et al. (46).

Excess relative risk

Figure 3 shows that UkrAm estimates drop much faster than estimates from Japanese a-bomb survivors (44). Such trends in the UkrAm cohort are likely to be influenced by screening effects. They seem to shorten the latency period of radiation-induced cases and increase the slope of the age dependence. Our point estimates of the ERR at age 27 yr in Table S8 are identical to those of ref. (7). Our confidence intervals are markedly narrower since they are derived from 115 PTCs compared to 61 PTCs in ref. (7). Our ERR estimates are similar to those of refs. (9,11,44) at an attained age of 20 yr. They drop fast according to age^{-4} whereas for older ages other studies (9,44) report a substantial ERR of about 1 Gy^{-1} even decades after exposure below age 10 yr. The 95% CI of our ERR estimates include the point estimates from pooled studies of external exposure (3,46).

Retrospective analysis

Applying the relation

$$\text{ERR}(D) = P_{\text{mol}}(D)/(1 - P_{\text{mol}}(D)),$$

which links the radiation risk to the CLIP2 dose response from molecular measurements for young patients (AaO < 20 yr) (16), yields an ERR estimate of 5.2 (95% CI 2.1; 13) at 1 Gy. The relation is obtained by inverting Eq. (S19) for the probability of causation. The point estimates of the ERR from our preferred descriptive and mechanistic models coincide with the upper CI of the 'molecular' CLIP2 estimate, and all three lower CI are almost equal. The point estimate from CLIP2 measurements is equal to the epidemiological estimate of Tronko et al. (6) but CIs are markedly smaller (Figure 3).

Analysis of molecular measurements

To assess the impact of radiation on CNAs, transcriptomic profiles, BRAFV600 mutations and RET/PTC rearrangements we have estimated the share of path-specific cases and corresponding odds ratios in patients groups which were defined by the appearance of such molecular changes (18). A detailed breakdown of the analysis which addresses each molecular alteration separately is given in the Supplementary Material (Tables S9 and S10). Due to small case numbers the statistical analysis can only represent trends without claiming significance.

Copy number alterations

CNAs were investigated on 84 patients of the UkrAm cohort ((18), their Figure 2). In the present study we could only include 79 patients due to a lack of TDose10 estimates for the rest. Four CNA clusters based on similar aberration patterns were detected after a hierarchical analysis. Averaged patient data for these four clusters are given in Table S9. 22 cases from clusters 1.2 and 2.2 showed many CNAs. This observation coincides well with the model result of 24 cases from MSC underpinning the notion that many CNAs signify a hallmark of sporadic PTC. In clusters 1.2 and 2.2

the share of MSC cases is above the total average whereas the share of radiation-induced cases lies below. Corresponding to this observation radiation-induced cases are overrepresented in clusters 1.1 and 2.1 where CNAs are almost absent. This is also a common feature of PTC patients below age 35 yr from the TCGA (The Cancer Genome Atlas) data base (cancergenome.nih.gov) (18,47). Since TCGA patients have no reported radiation history the absence of CNAs is not related to radiation in the UkrAm cases.

Transcriptome expression

mRNA expression data from UkrAm cases published by Abend et al. (48) were investigated for molecular subtypes of 'RAS-like' and 'BRAF-like' PTCs that have been previously published by the TCGA consortium (47). The UkrAm cases were divided into four clusters C1 – C4 by applying a 59-gene signature (18) and revealed a 'RAS-like', an 'intermediate RAS-BRAF', a 'BRAF-like' and a so far unknown subgroup (C4) that might have represented radiation-associated PTCs (Table S10) (18). This assumption is supported by the present analysis because the mechanistic model predicts that six out of nine cases with complete patient data from cluster C4 are radiation-induced.

Gene mutations and rearrangements

Moreover, only three out of eight cases harboring BRAF mutations in the 'BRAF-like' cluster are related to radiation according to the mechanistic model (Table S10). The odds ratio of 0.65 further corresponds to a decreasing association of BRAF mutations with an increasing thyroid dose in a logistic regression analysis of the data set from ref. (18). Regression of RET/PTC rearrangements on the thyroid dose did not reveal an association suggesting an equal occurrence of this characteristic alteration in radiation-induced and sporadic PTCs (18) again in line with model predictions.

Discussion

The main goal of our study was to harness biomarker information for narrowing the gap between molecular biology and radiation epidemiology. Already descriptive risk assessment benefits from the introduction of such a marker. Good accordance of the CLIP2 status determined by model predictions and measurements in Tables 2 and S7 indicates carcinogenic effects of radiation already at doses below 100 mGy. For the first time a 'molecular' ERR has been derived solely from the dose response of marker measurements by Selmansberger et al. (16) in agreement with the epidemiological result of Tronko et al. (6) (Figure 3). The PTC cases in both studies were mostly operated before the year 2001 at a similar mean age of about 17 yr.

For mechanistic risk assessment we present a novel way to integrate molecular genetic data into a model of carcinogenesis. Our multi-scale approach spans several size dimensions from point mutations to tumors. The model design is based on the property of CLIP2 as a binary radiation marker (14,15). It features the progression of cell-based key events on the path to cancer (21,22). Development of PTC is considered separately in a CLIP2-associated path (C2C) and a CLIP2-independent path (MSC) (Figure 1).

True to its dual purpose the model is validated with both molecular measurements and epidemiological data. It reproduces the observed dose response of CLIP2 overexpression in young patients (Figure 2) (16). Based on this response, model predictions on the frequency of CLIP2-positive/negative cases conform remarkably to the measured CLIP2 data from 141 PTC patients contained in the Chernobyl tissue bank (Tables 2 and S7). Risk estimates are in good agreement with age and dose-dependent estimates for thyroid cancer after Chernobyl in cohorts UkrAm and BelAm, and with estimates from pooled studies of external childhood exposure (3,6-8,46) (Figure 3).

Selmansberger et al. (15) postulated CLIP2-negative markers exclusively in unexposed patients and CLIP2-positive markers in exposed patients. Under these assumptions a marker sensitivity of 76%

and a marker specificity of 75% were found. However, the present model predictions reveal a more complex situation in Table 2. CLIP2-positive cases may appear in unexposed patients and CLIP2-negative cases may appear in exposed patients. Underpinned by the close agreement between measured and predicted cases we expect a much higher reliability of the CLIP2 marker with both specificity and sensitivity closer to the optimal value of 100%. Yet more accurate estimates for both quantities are difficult to obtain. Since the model can only assign a probability for the CLIP2 status of a single PTC case the 'true' subdivision of cases according to their CLIP2 status is not available. After cross-checking with both model predictions of the present study and an independent epidemiological risk estimate (6) CLIP2 emerges as a radiation marker of high quality.

With the validated model various forms of molecular changes in PTC were analyzed to obtain a more comprehensive characterization of the molecular landscape. The number of PTC cases with many CNAs is nearly reproduced by the model suggesting that CNAs mainly arise in the sporadic MSC path (Table S9). The cumulative occurrence of CNAs on chromosomes 1p and 19 ((18), their Figure 2) might be related to the two or three driver gene mutations which are assumed in MSC (Figure 1). The model design is further supported by the rare occurrence of genetic alterations for radiation-induced PTC via the CLIP2-associated path as postulated by Takano (39) (Table S9). Radiation effects on BRAFV600E mutations, RET/PTC rearrangements and transcriptomic changes were also investigated with the model. Its predictions look plausible by comparison with published data on UkrAm cases (Table S10) (18).

Williams (49) proposed an alternative model concept which starts with thyrocytes as exclusive cells of origin for PTC. It relies on similar initial molecular changes possibly related to RET/PTC rearrangements for both sporadic and radiation-induced PTC. These early mutations are then amplified during childhood by growing cell clones in the follicular epithelium. Terminal mutations in initiated clones suppress growth limitations in young adults and give rise to well differentiated PTC later on. For radiation-induced PTC this concept is not supported by our analysis which suggests a

distinct path of CLIP2-associated carcinogenesis with its own molecular pattern and temporal development. RET/PTC rearrangements in particular do not show a dose response (18) (Table 3). A single path model with a dose response in initial mutations was designed with the concept of Williams (49). Fits of this model with eight parameters to the epidemiological data yielded a deviance about 40 points higher compared to the preferred mechanistic model.

We were able to relate different paths of PTC development to both molecular changes (Tables S9, S10) and epidemiological covariables (Table S11). Table 3 summarizes the characteristic features of PTC development in young Ukrainian patients exposed below age 19 yr after Chernobyl or born after the accident. However, a one-to-one relation of molecular changes to model paths is not entirely supported by measurements (18). The lack of full concordance reveals a weakness of the model which relies on a rigid path-specific classification. Yet in measured data a certain degree of overlap appeared.

Other limitations include the method of case ascertainment. Age at operation served as surrogate for age at onset of cancer but for some patients surgery was delayed by more than a year. We partially accounted for this uncertainty in the construction of the likelihood function which adapts prevalent and incidental cases differently. The date of fine needle aspiration for a confirmed PTC case would constitute a more accurate surrogate. But this information was not available for our study. The impact of dose uncertainties on risk estimates has not been taken into account. We expect small changes from corrected doses in the order of ten percent which have been found by Little et al. (50) for the cohort of prevalent cases in ref. (6). It remains an open question if the present modelling approach represents general features of PTC development. This should be tested by applying the model to other cohorts such as the Japanese a-bomb survivors.

By leaving an imprint in the incidence data, CLIP2 alterations constitute a rate-limiting event.

Moreover, recent analysis of the CLIP2 interaction network based on expression data from UkrAm cases suggests its involvement in fundamental oncogenic processes such as MAPK signaling,

apoptosis and chromosome segregation (18). Altogether, the findings from mechanistic modelling of epidemiological data and from molecular biology point to a function of CLIP2 as a driver gene in radiation-induced PTC. This result is new with respect to the recently published TCGA study which did not include radiation-associated PTCs (47). But for a final answer dedicated cell culture models are needed to demonstrate the oncogenic potential of CLIP2 overexpression.

Accepted Manuscript

Glossary

PTC	papillary thyroid cancer
CTB	Chernobyl tissue bank, www.chernobyltissuebank.com
CNA	copy number alteration
MSC	multi-stage carcinogenesis
C2C	CLIP2-related carcinogenesis
sC2C	sporadic CLIP2-related carcinogenesis
rCSC	radiation-induced CLIP2-related carcinogenesis
P_{mol}	probability of causation from CLIP2 measurements of ref. (16)
P_{epi}	probability of causation from the present epidemiological analysis, Eq. (S19)
AaO	age at operation (or surgery) (yr)
AaE	age at exposure (yr)
TsE	time since exposure (yr)

Funding

This work was supported by the European Commission under FP7 project EpiRadBio (Fission-2010-3.1.1, project no. 269553).

Acknowledgements

We thank Alina Brenner for discussions on the oblast-specific correction of risk estimates.

Conflicts of interest statement: None declared

References

1. de Gonzalez, A.B., *et al.* (2016) Relationship between paediatric CT scans and subsequent risk of leukaemia and brain tumours: assessment of the impact of underlying conditions. *Br. J. Cancer*, **114**, 388-94.
2. Eidemüller, M., *et al.* (2015) Breast cancer risk and possible mechanisms of radiation-induced genomic instability in the Swedish hemangioma cohort after reanalyzed dosimetry. *Mutat. Res.*, **775**, 1-9.
3. Veiga, L.H.S., *et al.* (2016) Thyroid Cancer after Childhood Exposure to External Radiation: An Updated Pooled Analysis of 12 Studies. *Radiat. Res.*, **185**, 473-84.
4. Cardis, E., *et al.* (2011) The Chernobyl accident-an epidemiological perspective. *Clin Oncol (R Coll Radiol)*, **23**, 251-60.
5. Wakeford, R. (2011) The silver anniversary of the Chernobyl accident. Where are we now? *J Radiol Prot*, **31**, 1-7.
6. Tronko, M.D., *et al.* (2006) A cohort study of thyroid cancer and other thyroid diseases after the chornobyl accident: thyroid cancer in Ukraine detected during first screening. *J. Natl. Cancer Inst.*, **98**, 897-903.
7. Brenner, A.V., *et al.* (2011) I-131 dose response for incident thyroid cancers in Ukraine related to the Chornobyl accident. *Environ. Health Perspect.*, **119**, 933-9.
8. Zablotska, L.B., *et al.* (2011) Thyroid cancer risk in Belarus among children and adolescents exposed to radioiodine after the Chornobyl accident. *Br. J. Cancer*, **104**, 181-7.
9. Furukawa, K., *et al.* (2013) Long-term trend of thyroid cancer risk among Japanese atomic-bomb survivors: 60 years after exposure. *International Journal of Cancer*, **132**, 1222-6.
10. Blettner, M. (2015) Comment: The merits and limits of pooling data from nuclear power worker studies. *The Lancet Haematology*, **2**, e268 - e269.
11. Preston, D.L., *et al.* (2007) Solid cancer incidence in atomic bomb survivors: 1958-1998. *Radiat Res*, **168**, 1-64.
12. Pernot, E., *et al.* (2012) Ionizing radiation biomarkers for potential use in epidemiological studies. *Mutat. Res.*, **751**, 258-86.
13. Hall, E.J., *et al.* (2012) *Radiobiology for the Radiologist*. Lippincott Williams & Wilkins.
14. Hess, J., *et al.* (2011) 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.
15. Selmansberger, M., *et al.* (2015) CLIP2 as radiation biomarker in papillary thyroid carcinoma. *Oncogene*, **34**, 3917-25.
16. Selmansberger, M., *et al.* (2015) Dose-dependent expression of CLIP2 in post-Chernobyl papillary thyroid carcinomas. *Carcinogenesis*, **36**, 748-56.
17. Preston, R.J. (2015) Integrating basic radiobiological science and epidemiological studies: why and how. *Health Physics*, **108**, 125-30.
18. Selmansberger, M., *et al.* (2015) Genomic copy number analysis of Chernobyl papillary thyroid carcinoma in the Ukrainian-American Cohort. *Carcinogenesis*, **36**, 1381-7.
19. Moolgavkar, S.H. (1991) *Carcinogenesis models: An overview*. Plenum Press.
20. Little, M.P., *et al.* (2008) A stochastic carcinogenesis model incorporating multiple types of genomic instability fitted to colon cancer data. *Journal of Theoretical Biology*, **254**, 229-238.
21. Tomasetti, C., *et al.* (2015) Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 118-23.
22. Tomasetti, C., *et al.* (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**, 78-81.
23. Kaiser, J.C., *et al.* (2014) Genomic instability and radiation risk in molecular pathways to colon cancer. *PLoS ONE*, **9**, e111024.

24. Likhtarev, I., *et al.* (2003) Uncertainties in thyroid dose reconstruction after Chernobyl. *Radiation Protection and Dosimetry*, **105**, 601-8.
25. Likhtarov, I., *et al.* (2005) Post-Chornobyl thyroid cancers in Ukraine. Report 1: Estimation of thyroid doses. *Radiation Research*, **163**, 125-136.
26. Likhtarev, I., *et al.* (2006) Questionnaire- and measurement-based individual thyroid doses in Ukraine resulting from the Chornobyl nuclear reactor accident. *Radiation Research*, **166**, 271-86.
27. Likhtarov, I., *et al.* (2013) Estimating thyroid masses for children, infants, and fetuses in Ukraine exposed to ¹³¹I from the Chernobyl accident. *Health Physics*, **104**, 78-86.
28. Bianconi, E., *et al.* (2013) An estimation of the number of cells in the human body. *Ann. Hum. Biol.*, **40**, 463-71.
29. Dumont, J.E., *et al.* (1992) Physiological and pathological regulation of thyroid cell proliferation and differentiation by thyrotropin and other factors. *Physiol. Rev.*, **72**, 667-97.
30. Fierabracci, A. (2012) Identifying thyroid stem/progenitor cells: advances and limitations. *J. Endocrinol.*, **213**, 1-13.
31. Coclet, J., *et al.* (1989) Cell population kinetics in dog and human adult thyroid. *Clin. Endocrinol. (Oxf)*, **31**, 655-65.
32. Xing, M. (2013) Molecular pathogenesis and mechanisms of thyroid cancer. *Nat. Rev. Cancer*, **13**, 184-99.
33. Hamatani, K., *et al.* (2008) RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose. *Cancer Research*, **68**, 7176-82.
34. Leeman-Neill, R.J., *et al.* (2013) RET/PTC and PAX8/PPAR γ chromosomal rearrangements in post-Chernobyl thyroid cancer and their association with iodine-131 radiation dose and other characteristics. *Cancer*, **119**, 1792-9.
35. Su, X., *et al.* (2016) Radiation exposure, young age, and female gender are associated with high prevalence of RET/PTC1 and RET/PTC3 in papillary thyroid cancer: a meta-analysis. *Oncotarget*, **7**, 16716-30.
36. Rosenbaum, E., *et al.* (2005) Mutational activation of BRAF is not a major event in sporadic childhood papillary thyroid carcinoma. *Mod. Pathol.*, **18**, 898-902.
37. Unger, K., *et al.* (2008) Array CGH demonstrates characteristic aberration signatures in human papillary thyroid carcinomas governed by RET/PTC. *Oncogene*, **27**, 4592-602.
38. Zitzelsberger, H., *et al.* (2010) Molecular rearrangements in papillary thyroid carcinomas. *Clin. Chim. Acta*, **411**, 301-8.
39. Takano, T. (2007) Fetal cell carcinogenesis of the thyroid: theory and practice. *Semin. Cancer Biol.*, **17**, 233-40.
40. Luebeck, E.G., *et al.* (2013) Impact of tumor progression on cancer incidence curves. *Cancer Research*, **73**, 1086-96.
41. Meza, R., *et al.* (2015) Multistage carcinogenesis and the incidence of thyroid cancer in the US by sex, race, stage and histology. *BMC Public Health*, **15**, 789.
42. Kaiser, J.C. (2010) MECAN - A Software Package to Estimate Health Risks in Radiation Epidemiology with Multi-Model Inference, User's Guide.
43. R Core Team (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria.
44. Jacob, P., *et al.* (2014) Ultrasonography survey and thyroid cancer in the Fukushima Prefecture. *Radiation and Environmental Biophysics*, **53**, 391-401.
45. Jacob, P., *et al.* (2006) Thyroid cancer risk in areas of Ukraine and Belarus affected by the Chernobyl accident. *Radiat. Res.*, **165**, 1-8.
46. Ron, E., *et al.* (2012) Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. 1995. *Radiat. Res.*, **178**, AV43-60.

47. Agrawal, N., *et al.* (2014) Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell*, **159**, 676-690.
48. Abend, M., *et al.* (2012) Iodine-131 dose dependent gene expression in thyroid cancers and corresponding normal tissues following the Chernobyl accident. *PLoS One*, **7**, e39103.
49. Williams, D. (2015) Thyroid Growth and Cancer. *Eur Thyroid J*, **4**, 164-73.
50. Little, M.P., *et al.* (2014) Impact of uncertainties in exposure assessment on estimates of thyroid cancer risk among Ukrainian children and adolescents exposed from the Chernobyl accident. *PLoS ONE*, **9**, e85723.

Accepted Manuscript

Tables

Table 1. Number of PTC cases with CLIP2 status in the tumor cohort of the present study, which was formed by Selmansberger et al. (16) (see their online supplementary table 1), number of PTC cases in the epidemiological UkrAm cohort of the present study (cases without CLIP2 typing in brackets) and number PTC cases with CLIP2 typing; which belong to both cohorts, divided in groups of age at operation (AaO) before and after 20 yr.

AaO period	tumor cohort		cases common in both cohorts	epidemiological UkrAm cohort
	mean AaO (yr)	cases		
AaO \geq 20 yr	27.7	65	56 ^a	86 (30)
AaO < 20 yr	16.6	76 ^b	14 ^c	29 (15)
all AaO	21.7	141	70	115 (45)

^a3 patients with AaO \geq 20 yr were members of the UkrAm cohort according to information from the CTB; but were not present in the epidemiological UkrAm cohort of the present study

^bincluding 24 patients with nominal doses of 1 mGy/yr born after 1986 from Genrisk-T cohort

^c1 patient with AaO < 20 yr from cohort Genrisk-T was member of the epidemiological UkrAm cohort

Table 2. Number of PTCs with positive CLIP2 marker among the 141 PTCs in our tumor cohort broken down by cohorts Genrisk-T, Genrisk-T plus and UkrAm (16), measurement results are compared with mechanistic model predictions, mean values of age at exposure (AaE), time since exposure (TsE), age at operation (AaO) and thyroid dose are given where applicable.

						PTC cases with positive CLIP2 marker		
		arithmetic means				measured	predicted	
	PTC	AaE	TsE	AaO	dose		total	radiation-induced
cohort	cases	(yr)	(yr)	(yr)	(Gy)			
Genrisk-T								
unexposed ^a	17	-	-	14.9	0.015	3	5.0	4.9
exposed	15	1.9	15.9	17.8	0.278	11	11.1	10.6
Genrisk-T plus								
unexposed ^a	7	-	-	15.3	0.015	3	2.0	1.9
exposed	30	2.2	15.5	17.8	0.750	23	23.0	22.2
UkrAm patients								
AaO < 20 yr	13	2.2	15.6	17.8	1.558	10	11.5	11.4
UkrAm patients								
AaO ≥ 20 yr	59	9.4	17.9	27.3	1.175	43	39.0	25.4
total	141	-	-	17.0	0.577	93	91.5	76.4

^aunexposed patients are born after 1986 and are given a nominal dose of 1 mGy/yr

Table 3. Characterization of CLIP2-associated carcinogenesis (C2C) and multi-stage carcinogenesis (MSC) in papillary thyroid cancer by epidemiological covariables and molecular changes, odds ratios (OR as ratio of case shares with and without a given property) are calculated by the mechanistic model.

	CLIP2-associated carcinogenesis		multi-stage carcinogenesis	remarks
	radiation-induced	sporadic	sporadic	
	rC2C	sC2C	MSC	
CLIP2 marker	strong dose response	no dose response	absent	model concept ^a , Figure 1
susceptible cells	thyroblasts or thyrocytes ^b		thyrocytes	cell no. N_b , N_s , Figures S1, S2
pre-neoplastic lesions	absent	absent	(probably) present	clonal growth γ , Figure S2
share of cases ^c	54% (76 cases)	11% (15 cases)	35% (50 cases)	model result, Eq. S21
share of cases ^d	56% (64 cases)	16% (19 cases)	28% (32 cases)	model result, Eq. S21
age at exposure (yr)	< 5 (OR 1.8)	≥ 5 (OR 8.1)	≥ 5 (OR 1.6)	Table S11
time since exposure (yr)	< 17 (OR 1.4)	≥ 17 (OR 1.8)	≥ 17 (OR 1.5)	Table S11
age at operation (yr)	< 20 (OR 1.4)	≥ 20 (OR 30)	unspecific	Table S11
thyroid dose (Gy)	≥ 0.2 (OR 2.5)	< 0.2 (OR 1.8)	< 0.2 (OR 3.3)	Table S11
copy number alterations	few	NA ^e	many	Table S9
transcriptomic cluster	C4 ^f	NA ^e	C1, C2, C3 ^g	Table S10, see also ref. (47)
BRAF mutations	rare	NA ^e	frequent	Table S10, see also ref. (35)
RET/PTC rearrangem.	present	NA ^e	present	Table S10, see also refs. (33-35)

^amotivated by Selmansberger et al. (16)

^bhypothesis of Takano (39)

^c141 PTC cases from Selmansberger et al. (16)

^d115 PTC cases from UkrAm cohort in present study

^enot calculated due to expected small case numbers

^fC4 in Selmansberger et al. (18) (their online Figure S1), novel transcriptomic profile radiation-induced (model result)

^gC1, C2, C3 in Selmansberger et al. (18): 'RAS-like' (C1), 'intermediate RAS-BRAF' (C2), 'BRAF-like' (C3)

Figures and Legends

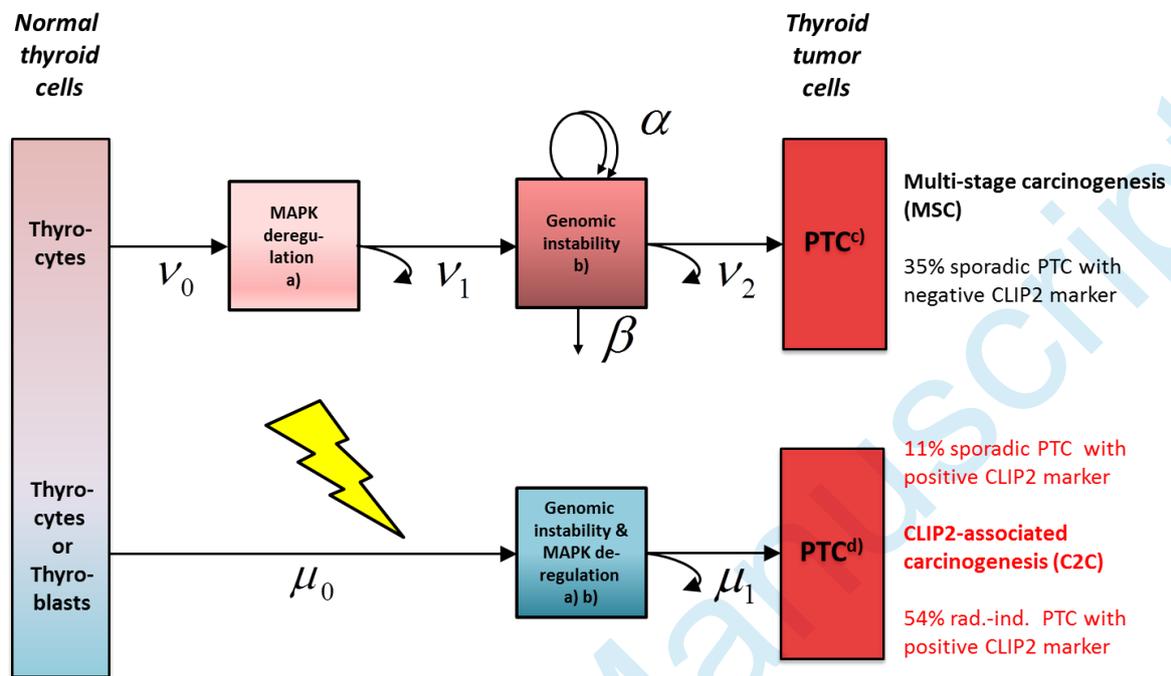


Figure 1. Conceptual model for development of sporadic PTC from multi-stage carcinogenesis (MSC) (upper path) and sporadic or radiation-induced PTC from CLIP2-associated carcinogenesis (C2C) (lower path), path-specific case shares are calculated from the preferred mechanistic model based on 141 PTC patients used in Selmansberger et al. (16), a mathematical implementation of the model is given in the Supplementary Material, black arrows with Greek symbols denote transition rates between boxes, jagged yellow arrow denotes radiation action on rate, boxes represent cells with molecular changes discussed in Selmansberger et al. (18) for a) point mutations of genes BRAF, RAS or gene rearrangements (e.g. RET/PTC) (Table S10), b) copy number alterations (MSC) or CLIP2-associated aneuploidy (Table S9), c) transcriptomic subtypes 'RAS-like', 'intermediate RAS-BRAF', 'BRAF-like' (Table S10), d) novel transcriptomic subtype (radiation-related) (Table S10).

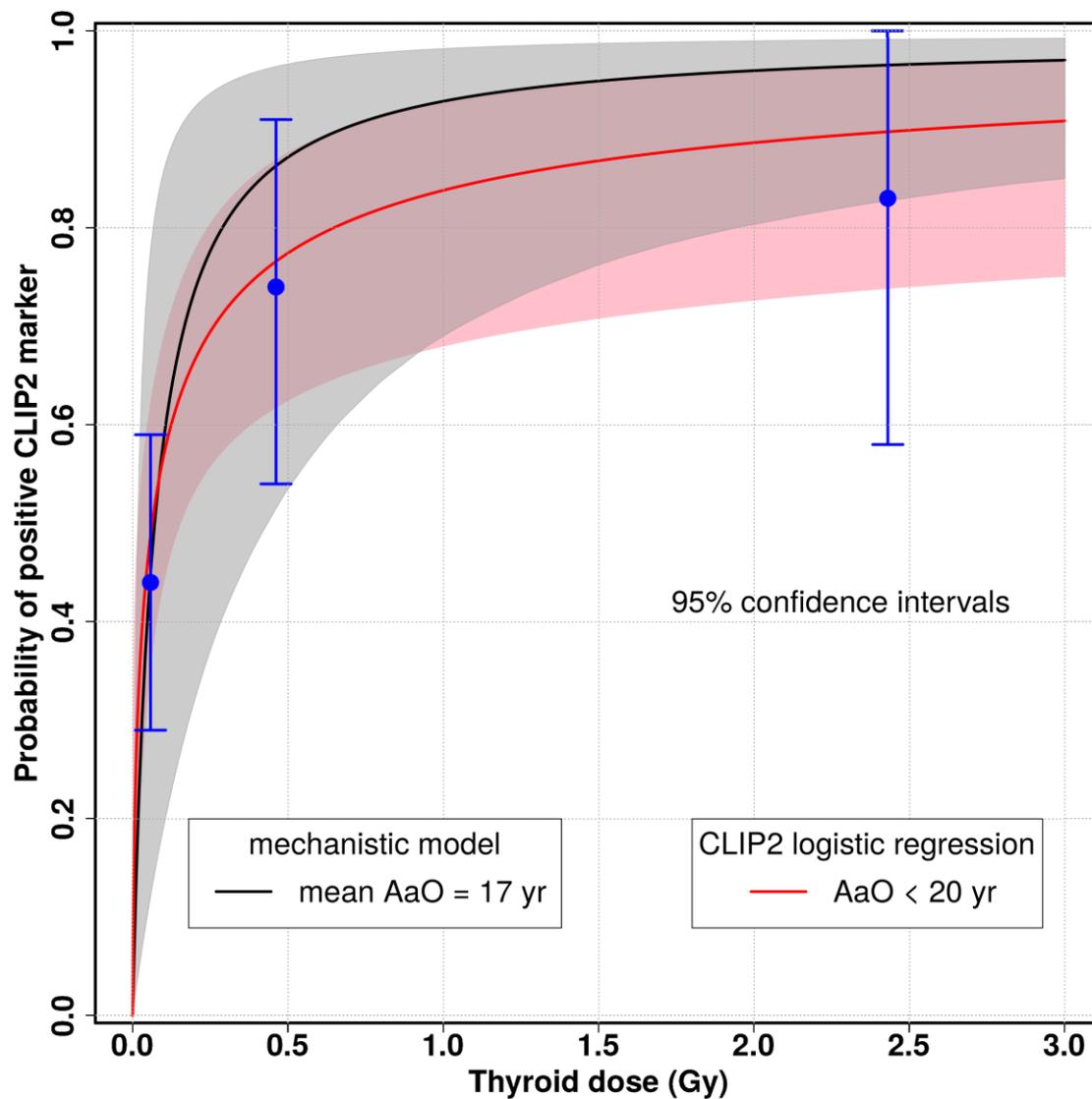


Figure 2. Dose response for the probability $P(C2C|PTC)$ of detecting a positive CLIP2 biomarker in a PTC case at attained age 17 yr (mean age in group AaO < 20 yr) from mechanistic model (black line) compared to the dose response for from logistic regression on molecular measurements for AaO < 20 yr (16) (red line); crude rates (blue points) from Table S8.

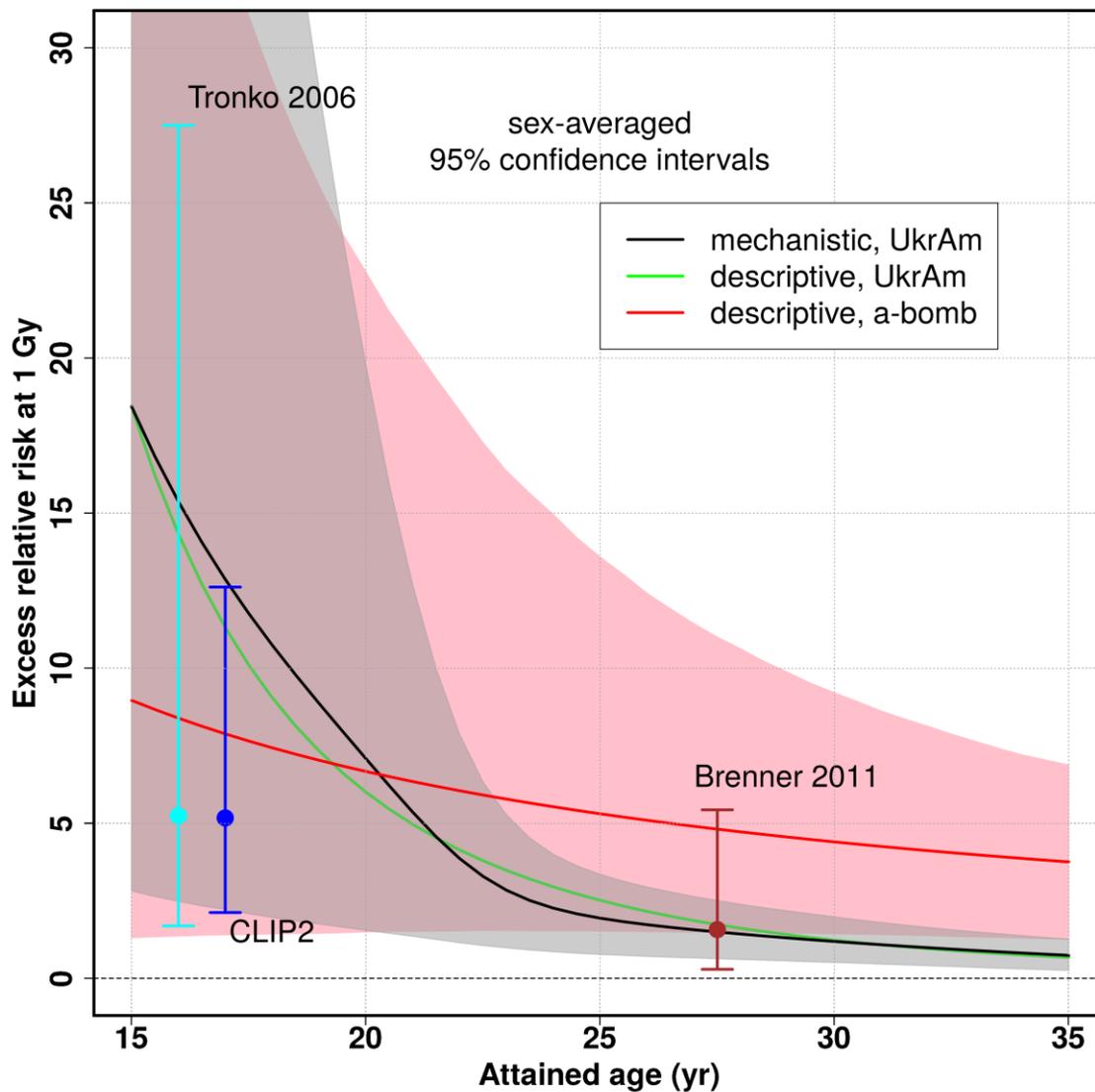


Figure 3. Age dependence of the sex-averaged ERR at a thyroid dose of 1 Gy for the descriptive model (green line) and the mechanistic model (black line with gray-shaded 95% CI); for comparison the estimates of Tronko et al. (6) (all histological types for 1st screening), Brenner et al. (7) (PTC only for 2nd – 4th screening), from CLIP2 measurements in PTCs (AaO < 20 yr) and of Jacob et al. (44) (all histological types) for Japanese a-bomb survivors exposed at age 8 yr (red line with pink 95% CI) are shown.