February 3, 2012

CHILDHOOD CANCER SURVIVOR STUDY
ANALYSIS CONCEPT PROPOSAL

I. Title: Genome-Wide Association Study of Subsequent Neoplasms Among Childhood Cancer Survivors

II. CCSS Working groups: Genetics (primary), Second Malignancies (secondary), and Epidemiology/Biostatistics (secondary)

Investigators, by primary area of expertise:

Epidemiology

Lindsay M. Morton, PhD, National Cancer Institute (mortonli@mail.nih.gov)
Preetha Rajaraman, PhD, National Cancer Institute (rajarama@mail.nih.gov)
Smita Bhatia, MD, MPH, City of Hope Comprehensive Cancer Center (sbhatia@coh.org)
Leslie L. Robison, PhD, St. Jude Children's Research Hospital (les.robison@stjude.org)
Joseph P. Neglia, MD, MPH, University of Minnesota (jneglia@tc.umn.edu)
Gregory Armstrong, MD, M.S.C.E., St. Jude Children's Research Hospital (greg.armstrong@stjude.org)
Amy Berrington de Gonzalez, PhD, National Cancer Institute (berringtona@mail.nih.gov)
Joseph F. Fraumeni, Jr., MD, National Cancer Institute (fraumeni@mail.nih.gov)

Genetics

Stephen J. Chanock, MD, National Cancer Institute (chanocks@mail.nih.gov)
Margaret A. Tucker, MD, National Cancer Institute (tuckerp@mail.nih.gov)
Louise Strong, PhD, MD Anderson Cancer Center (lstrong@mdanderson.org)

Biostatistics

Joshua N. Sampson, PhD, National Cancer Institute (sampsonjn@mail.nih.gov)
Wendy M. Leisenring, ScD, Fred Hutchinson Cancer Research Center (wleisenr@fhcrc.org)
Yutaka Yasui, PhD, University of Alberta (yyasuiua@gmail.com)

Dosimetry

Marilyn Stovall, PhD, M.D. Anderson Cancer Center (mstovall@mdanderson.org)
Susan A. Smith, MPH, M.D. Anderson Cancer Center (sasmith@mdanderson.org)
Rita E. Weathers, MS, M.D. Anderson Cancer Center (rweather@mdanderson.org)
Irene Harris, BS, CMD, M.D. Anderson Cancer Center (iharris@mdanderson.org)

III. Background and rationale:

Major advances in cancer treatments have led to dramatic improvements in the five-year relative survival following a childhood cancer, from approximately 60% in the 1970s to over 80% today.[1] Subsequent neoplasms are a leading cause of morbidity and mortality among childhood cancer survivors.[2-5] With 20% of childhood cancer survivors developing a subsequent neoplasm within 30 years following diagnosis,[4,5] identifying patients at highest risk for this potentially fatal outcome is of paramount importance.

Prior cancer treatments are one of the strongest risk factors for subsequent neoplasms in childhood cancer survivors. Radiotherapy is commonly used to treat childhood cancer, yet ionizing radiation is an established carcinogen for a number of neoplasms, with higher relative risks often seen for individuals exposed at younger ages.[6-13] Previous studies of childhood cancer survivors have demonstrated strong dose-response relationships for radiation dose from radiotherapy with risk of subsequent breast cancer,[14,15] thyroid cancer,[8,12] central nervous system (CNS) neoplasms,[9] and bone and soft-tissue sarcomas.[16,17] In addition, ionizing radiation exposure increases risk for non-melanoma skin cancer

(NMSC), particularly basal cell carcinoma,[18,19] and acute leukemia.[6,20] Together, these highly radiosensitive neoplasms account for approximately three-quarters of subsequent neoplasms occurring within several decades of a childhood cancer.[5,21]

Many childhood cancer patients also are treated with chemotherapy, alone or in combination with radiotherapy. Cytotoxic chemotherapies, such as alkylating agents and topoisomerase II inhibitors, have long been recognized as leukemogenic.[22] However, chemotherapy is increasingly recognized as contributing to non-hematologic malignancies as well (e.g., thyroid, sarcoma, lung, stomach, bladder), with increased risks observed for other classes of chemotherapeutic drugs such as anthracyclines and antimetabolites in addition to alkylating agents, even after adjustment for radiation dose.[5,14-17,23-25]

Individuals with certain hereditary disorders such as ataxia telangiectasia have marked sensitivity to the effects of radiation, but less is known about genetic susceptibility to radiation-related carcinogenesis beyond the context of these rare disorders,[26] and very little is known about genetic susceptibility to chemotherapy-related carcinogenesis.[27] It is likely that multiple, complex genetic pathways such as DNA damage repair, oxidative stress, and cell cycle control contribute to the development of radiation- and chemotherapy-related neoplasms, supporting a polygenic model for sensitivity to therapy-related neoplasms. Such inherited sensitivity is thought to play a greater role in the development of cancer in children compared with adults, as evidenced by the early age at cancer onset that characterizes many of the known familial cancer predisposition syndromes [e.g., Li-Fraumeni syndrome (*TP53*),[28] Cowden syndrome (*PTEN*),[29] and Lynch syndrome (DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, or *PMS2*)[30]].

The Childhood Cancer Survivor Study (CCSS) provides a unique opportunity to study the risk factors for subsequent neoplasms following childhood cancer because of its long-term follow-up of over 14,000 childhood cancer survivors (of whom more than 1400 have developed at least one subsequent neoplasm to date), the availability of biospecimens for nearly half of participants, and high quality information on radiotherapy and chemotherapy exposures.[31] We therefore propose to conduct a genome-wide association study (GWAS), an agnostic evaluation of genetic markers [tag single nucleotide polymorphisms (SNPs)] across the genome, within CCSS to identify heretofore unsuspected regions of the genome that may predispose for the development of cancer, with a focus on therapy-related neoplasms. The promise of this approach was recently demonstrated in the CCSS GWAS of second malignancies in Hodgkin lymphoma survivors treated with radiotherapy.[32] We seek to expand that analysis to study genetic susceptibility in CCSS participants across all first primary cancer types and treatments, with greater density of genetic markers and more detailed characterization of radiation and chemotherapy exposure.

This study will further our understanding of cancer susceptibility and the etiology of multiple primary cancers in childhood cancer survivors as well as elucidate potential mechanisms of radiation- and chemotherapy-related carcinogenesis. In addition, identification of patients with genetic susceptibility to multiple primary cancers has important translational potential for clinical decision-making for childhood cancer treatments (e.g., modifications of treatment regimens or doses) and/or follow-up plans (e.g., screening for early detection) to reduce the burden of subsequent neoplasms.

IV. Objective: Conduct a GWAS of subsequent neoplasms occurring after childhood cancer in CCSS.

Specific aims:

1) Identify genetic variants associated with the development of neoplasms subsequent to childhood cancer. We aim to identify and distinguish genetic variants that modify:
    A) the effect of radiotherapy on risk of subsequent highly-radiosensitive neoplasms (including cancers of the breast, CNS, and thyroid; sarcoma; acute leukemia; and NMSC),
    B) the effect of chemotherapy on the risk of subsequent neoplasms, and

C) the risk of subsequent neoplasms independent of treatment exposures.
2) Identify genetic variants associated with the risk of childhood cancer.
3) Develop a resource of genetic data that can be used by investigators to conduct secondary analyses of more specific hypotheses related to the aims listed above or to conduct analyses of other outcomes (e.g., cardiovascular events).

Research hypotheses:
We hypothesize that:
1) Children who develop multiple primary cancers have a broad inherited predisposition to cancer, enabling the identification of genetic regions that are major drivers of carcinogenesis in general and thus may be associated with multiple types of cancer.
2) Inherited genetic variation may alter the biological response of normal cells to DNA damage and immunosuppression from ionizing radiation and/or chemotherapeutic agents and thereby alter cancer susceptibility.
3) Inherited genetic variation also contributes to the spectrum of adverse outcomes[33] observed among childhood cancer survivors.

V. Analysis framework:

Outcomes of interest

The outcome of interest is the occurrence of subsequent neoplasms, including malignant (invasive and *in situ* cancers) and certain benign tumors (Table 1). For the analysis of radiotherapy-related subsequent neoplasms, the outcome will be restricted to neoplasms that have been shown to be highly-radiosensitive in this and other study populations (including cancers of the breast, CNS, and thyroid; sarcoma; acute leukemia; and NMSC) to increase statistical power to evaluate radiation-specific genetic variants.

Subject population

Patients eligible for CCSS were diagnosed before age 21 years with a specific childhood malignancy (leukemia, CNS tumor, Hodgkin lymphoma, non-Hodgkin lymphoma, renal tumor, neuroblastoma, soft-tissue sarcoma, or bone tumor) at one of 26 participating centers in the United States and Canada during 1970-1986 and were alive at least five years after their original diagnosis.[31] CCSS also recruited siblings of some patients for ancillary studies. For the present study, eligible patients must have a minimum amount of DNA (as described further below) and have at least some follow-up for the occurrence of subsequent neoplasms. For evaluation of therapy-related neoplasms, eligible patients also must have agreed to the release of their medical records to obtain information on radiotherapy and chemotherapy treatments.

Biospecimen availability

Biospecimen collection (blood or buccal cell, via mouthwash initially and using Oragene since 2004) began in the 1990s, several years after the CCSS cohort was assembled. Among individuals who did not provide a biospecimen initially, those who developed a subsequent neoplasm were targeted as a higher priority for requesting biospecimens, though efforts are ongoing to collect a biospecimen for all individuals. Biospecimens currently are available for approximately 6900 individuals in CCSS (48% of the total cohort, Table 2), including over 900 (67%) individuals who have developed at least one subsequent neoplasm by the most recent follow-up date ("cases") and nearly 6000 (46%) individuals who have not developed a subsequent neoplasm by the most recent follow-up date ("controls"). Table 3 provides the breakdown of biospecimen availability by primary diagnosis and subsequent neoplasm type.

For the present study, eligible individuals must have at least 500ng of available DNA from any source. Eligibility for individuals who received an allogeneic bone marrow transplant will be restricted to those with at least 500ng of available DNA from a buccal cell sample to prevent contamination with donor DNA in blood cells. For those cases who do not have at least 500ng of available DNA, we will substitute

a sibling when available (preliminary estimates suggest biospecimens from siblings will be available for approximately 40 cases who do not have a biospecimen).

A second control set will be derived from the Division of Cancer Epidemiology and Genetics (DCEG) Total Genome Set (TGS), a series of individuals with existing GWAS data from other studies who were known to be cancer-free as of age 55 years. A random sample of these individuals will be selected by stratified random sampling from the cohort, frequency matching by sex and race, with the aim of identifying about 10,000 individuals.

Key variables

Requested data include basic demographic information (e.g., sex, race/ethnicity, date of birth), information on all primary neoplasm diagnoses (including site, histology, date of diagnosis, and microscopic confirmation), and all available treatment data (for the first primary cancer as well as any subsequent neoplasms, recognizing that data were collected systematically only for those treatments occurring within 5 years of the first primary cancer). For subsequent neoplasms, any additional detailed information on tumor location is requested as well. Finally, data on body-mass index and hormonal factors are requested for consideration of potential confounding (but will not be used as outcomes).

Analyses that consider radiotherapy dose will include more detailed radiation dosimetry data (see below). Analyses focused on chemotherapy will consider broad classes of chemotherapy, including alkylating agent, anthracycline, antimetabolite, or epipodophyllotoxin-based chemotherapy, as well as individual drugs as allowed by sample size. Chemotherapy analyses that consider doses within a class of chemotherapeutic agents will use a scored variable that reflects the dose distributions across multiple agents within that class, as has been done in previous CCSS analyses and other studies of childhood cancer.[25,34]

Radiation dosimetry

Radiation doses will be estimated by collaborating medical physicists at M.D. Anderson Cancer Center using standard methodology.[35] Briefly, data on individual patients' radiotherapy fields and tumor dose already have been collected from radiotherapy and other medical records. These data form the foundation for estimating doses to specific locations in the body using a custom-designed dose program, based on measurements in water and anthropomorphic phantoms constructed of tissue-equivalent material.

For these analyses, patients will be categorized into high, medium, low, or no radiation dose (based on the maximum treatment dose) to specific body regions, including the brain (divided into 4 segments), thyroid, chest, abdomen, pelvis, arms, and legs. Analyses are underway to compare the categorization of patients using these estimated doses to a similar categorization approach based on doses derived from the detailed dosimetry conducted for patients selected for previous case-control studies, which estimated the mean dose to each case's tumor location (and corresponding locations in controls) and took patient-specific blocking into account. Although the case-control dosimetry is more precise in terms of both location and treatment details, it is not feasible to conduct such detailed dosimetry for each tumor location for the entire cohort. The planned comparisons will facilitate adjustment for measurement error in the analysis if needed or motivate development of an alternative approach to incorporating radiation dose into the analyses.

Genotyping approach

All samples will be received at the SAIC-Frederick DNA Extraction and Staging Lab and will undergo standard sample handling and evaluation procedures, including DNA quantitation and the Applied Biosystems Identifiler® assay. Genotyping will be conducted at NCI's Core Genotyping Facility.

A pilot study will be conducted to evaluate the quantity and quality of available DNA by specimen type, date of specimen collection, and "case/control" status. We propose to select 240 pilot specimens, including 120 from patients who have developed at least one subsequent neoplasm and 120 from patients who have not developed a subsequent neoplasm to date. Within each of these categories, the samples will be selected according to the following distribution: 10% blood collected in 2005 or later, 10% blood collected before 2005, 20% Oragene, 30% buccal collected in 2001 or later, and 30% buccal collected before 2001. Oragene samples were collected during a narrow time range (2008-2011) and thus are not broken down by date of collection. This distribution was chosen based on the assumption that the DNA quality is likely to be highest in the blood samples, next best in the Oragene samples, and lower in the mouthwash buccal cell samples. Table 2 provides the distribution of samples by primary cancer and specimen type.

The pilot aims to genotype N=192 individuals (must be a multiple of N=96). Thus, we propose to handle 25% above this number for a total of N=240 samples to increase the likelihood of 196 samples passing sample handling evaluations. Samples that pass sample handling evaluations but are not genotyped in the pilot will be genotyped in the full study.

DNA quality will be assessed by:

1. Completion of at least 13 markers on the Applied Biosystems Identifiler® assay. This assay performs multiplex PCR (16 amplicons) in a single tube, which requires high-quality double-stranded DNA (though the DNA may be fragmented). Availability of good data for at least 13 of these 16 markers is correlated with adequate performance on Illumina GWAS chips in the experience of NCI's Core Genotyping Facility (CGF).
2. SNP completion rate on the Illumina® HumanOmni5-Quad BeadChip. Note that only specimens that pass step 1 and have a sufficient amount of DNA will be genotyped.

For the full study, we propose to genotype all CCSS patients with adequate DNA quantity and quality at NCI's Core Genotyping Facility using the Illumina® HumanOmni5-Quad BeadChip, which covers minor allele frequencies (MAFs) as low as 1%, or similar technology (to be finalized at the time of genotyping due to rapid changes in genotyping costs and pending the level of available funds). Samples with sufficient DNA quality but insufficient quantity will be replenished. For cases with inadequate DNA, if a replacement sample is not available, a sibling will be substituted when available. Samples determined to have inadequate DNA for the GWAS will be reserved for replication. At least 75% of samples for each specimen type should pass sample handling evaluations to be included.

Cases and controls from replication studies (described below) will be genotyped by TaqMan® SNP Genotyping or similar technology (e.g., Fluidigm Dynamic Array) for the most noteworthy findings that emerge from the GWAS, based on a 1 d.f. p-value from the additive model in any analyses.

Analytic approach

Aim 1: The primary analysis will test for associations between each SNP and the diagnosis of a subsequent neoplasm using Cox proportional hazards regression for multiple outcomes,[36] adjusted for calendar year, primary diagnosis, chemotherapeutic agent, radiation dose, sex, and ethnicity (using standard principal component methods to adjust for ethnicity[37]). Each type of subsequent neoplasm will be treated as its own outcome, with radiation dose measured to the body region in which the subsequent neoplasm arose. Gene effects, including main effects and interactions with radiation and chemotherapeutic agents, will be allowed to differ by outcome through inclusion as random variables in a mixed model. The main objective of our primary analysis is to detect SNPs that are associated with the outcome either in the full population or only in a subgroup of individuals (e.g., those receiving chemotherapy, those receiving radiation, those receiving low dose radiation). This flexibility is permitted by testing whether the addition of SNP and SNP-by-treatment interaction terms, together, to our risk

model affect the probability of disease, rather than evaluating an "interaction" model of a particular form specified *a priori*. A secondary objective is to identify those SNPs that only influence risk of disease when receiving a specific type of treatment (e.g., those SNPs that alter the toxicity of chemotherapeutic agents). To make those identifications, we need a large enough sample of individuals receiving that treatment to detect that there is an effect on risk, and a large enough sample of individuals not receiving that treatment to detect that there is no more than a nominal effect. It is unlikely that we will have the statistical power in this study, by itself, to make those claims. However, there will be suggestive evidence as to which SNPs only affect risk in particular patient subgroups, and follow-up studies (e.g., replication, additional GWAS, or laboratory experiments) may offer more conclusive results.

Additional models will limit analyses to specific populations (e.g., Caucasians, certain subsequent neoplasm types, or survivors of a specific childhood cancer), exclude siblings, permit gender-specific effects, or allow for effect-size to be inversely related to radiation dose. Because longer-term survivors are more likely to have an available biospecimen, we will run appropriate sensitivity analyses to determine whether a SNP-survival association could be confounding any of our detected associations (though it is noted that detecting such a SNP would still prove valuable from the clinical perspective).

Aim 2: The TGS will be used as a control set for evaluating genetic susceptibility to childhood cancer using the standard case/control logistic regression approach.

Statistical power

Aim 1: We estimate the power for detecting an association between a SNP and the risk of a subsequent neoplasm. Although there are many scenarios, we consider a representative example where 50% of the individuals receive radiation therapy and assume that radiation increases the odds of each type of subsequent neoplasm by a factor proportional to $\exp(\beta_0 GR)$, where G is the number of minor alleles and R is radiation level (uniformly ranging from 0 to 1). If we consider an allele with MAF=0.10 and a 1:5 case:control ratio, we can expect to have 80% power to detect an association if each SNP increases the effect of radiation on odds of second cancer by a factor of 2.9. Table 4 shows estimates for power with other effect sizes and case:control ratios.

If we consider a more ideal scenario where 70% of the population receives a high dose of radiation and the SNP increases the risk of all types of subsequent neoplasms. Then, continuing with the hypothesis that the SNP only effects risk in radiotherapy-treated patients, we would now have 80% power to detect an OR=1.7 (Table 5). However, we could consider a more pessimistic scenario where 70% of the population receives radation, the radiation dose is uniformly distributed over its range, but the risk of only 50% of the subsequent neoplasm types are affected by the SNP. Here, we would have 80% power to detect an OR=4.2 (Table 6). Assuming a 1:5 case:control ratio, Figure 1 presents additional power estimates by effect size and number of cases. Details about the power calculations for Aim 1 are provided in the Appendix.

Aim 2: Assuming 10,000 appropriate matched individuals from the TGS for the case/control analysis, then, for SNPs with MAF=0.10, we will have 80% power to detect a SNP with a OR=3.3 in all childhood cancers that include more than 100 individuals (e.g., acute myeloid leukemia, Ewing sarcoma, and osteosarcoma), to detect a SNP with OR=1.8 in all childhood cancers that include more than 500 individuals (e.g., non-Hodgkin lymphoma, kidney tumors, and soft tissue sarcoma), and to detect a SNP with OR=1.6 in all childhood cancers that include more than 1000 individuals (e.g., acute lymphoblastic leukemia).

Replication

Aim 1: Childhood cancer survivors from the Key Adverse Events study of the Children's Oncology Group and the St. Jude Life Study cohort will serve as replication, totaling approximately 900 children

with at least one subsequent neoplasm and 1200 without a subsequent neoplasm. Patients also may be identified from the expanded CCSS cohort, including children diagnosed with a childhood cancer in the United States during 1987-1999. Finally, discussions have been initiated with key collaborators to identify other study populations for replication of our results, including a smaller study of childhood cancer survivors in France that has begun collecting biospecimens[38] and other studies of radiation-related neoplasms such as brain cancers.[39,40]

Aim 2: Replication for variants associated with childhood cancer will be derived from previously published scans of childhood cancers (e.g., acute lymphoblastic leukemia[41,42]), as well as ongoing scans of childhood cancers (e.g., Ewing's sarcoma, osteosarcoma, and Hodgkin lymphoma), noting that our study is restricted to 5-year survivors of childhood cancer.

Potential tables and figures

We anticipate publishing a primary manuscript focused on genetic regions associated with therapy-related neoplasms, with an emphasis on radiotherapy. Secondary manuscripts will focus on genetic regions associated with chemotherapy-related neoplasms, genetic regions that do not appear to be modified by radiotherapy or chemotherapy, and genetic regions associated with occurrence of multiple subsequent neoplasms. The tables and figures for each manuscript likely will be similar:

Tables

1. Selected characteristics of cases and controls from the Childhood Cancer Survivor Study and other study populations used in replication analyses (e.g., sex, race, first primary cancer diagnosis, calendar year of first primary cancer, age at first primary cancer, treatments for first primary cancer, diagnoses of subsequent neoplasms, time from first primary cancer to subsequent neoplasms).
2. SNPs identified in the GWAS of subsequent malignancies after childhood cancer (SNP, stage, number cases/controls, minor allele frequency, allelic odds ratio, 95% confidence interval, p-value).

Figures

1. Manhattan plot for the SNPs identified in the GWAS of subsequent neoplasms after childhood cancer, plotting the $-\log_{10}$P-value for each SNP against its respective position on each chromosome (including discovery, replication, and combined p-values).
2. Association results, recombination and linkage disequilibrium plots for specific chromosomal regions achieving genome-wide significance.

Genotype data sharing

In accordance with the NIH GWAS policy, a publication quality build of the dataset will be submitted to the database of Genotypes and Phenotypes (dbGaP) once the main effect manuscript is in press at a journal. Datasets managed within DCEG/NCI are designated as "CGEMS datasets". Access to controlled data is granted by the CGEMS Data Access Committee (DAC) of the NCI. Users requesting access to controlled data must submit a Data Access Request (DAR) to the CGEMS DAC for approval. DAC approval for controlled data access will be dependent upon completion of the DAR, agreeing to the terms and conditions in the Data Use Certification (DUC), and confirmation that the proposed research use is consistent with any restrictions on data use identified by the institutions that submitted the dataset to dbGaP. A biomedical research scientist from a recognized research institution can access both the genotype data and the executive summaries. All identifiers will be removed and only limited covariate data (case/control status, age group, and sex) will be available so as to prevent identification of subjects. Any other data (i.e., any other covariates) will only be accessible through the CCSS.

VI. Timeframe:

Required Approvals

September 2011	Senior Leadership in Genomics Committee (DCEG/NCI)
January 2012	Radiation Epidemiology Branch (DCEG/NCI)

	GWAS Certificate (St. Jude Institutional Review Board)
February 2012	(anticipated) CCSS Steering Committee
	(anticipated) Genotyping Review Committee (DCEG/NCI)
	(anticipated) Institutional Review Board (NCI)

Research Timeline

February-March 2012	Pilot study
April-May 2012	Sample receipt and handling for full study
June-July 2012	Genotyping of discovery set
July-December 2012	Analysis and replication

VII. References:

1. Howlader, N., *et al.* SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2008/. (2011).
2. Mertens, A.C., *et al.* Cause-specific late mortality among 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. *J Natl Cancer Inst* **100**, 1368-1379 (2008).
3. Reulen, R.C., *et al.* Long-term cause-specific mortality among survivors of childhood cancer. *JAMA* **304**, 172-179 (2010).
4. Armstrong, G.T., *et al.* Occurrence of multiple subsequent neoplasms in long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *J Clin Oncol* **29**, 3056-3064 (2011).
5. Friedman, D.L., *et al.* Subsequent neoplasms in 5-year survivors of childhood cancer: The Childhood Cancer Survivor Study. *J Natl Cancer Inst* **102**, 1083-1095 (2010).
6. Little, M.P., *et al.* Risks of leukemia in Japanese atomic bomb survivors, in women treated for cervical cancer, and in patients treated for ankylosing spondylitis. *Radiat Res* **152**, 280-292 (1999).
7. Preston, D.L., *et al.* Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. *Radiat Res* **158**, 220-235 (2002).
8. Sigurdson, A.J., *et al.* Primary thyroid cancer after a first tumour in childhood (the Childhood Cancer Survivor Study): a nested case-control study. *Lancet* **365**, 2014-2023 (2005).
9. Neglia, J.P., *et al.* New primary neoplasms of the central nervous system in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *J Natl Cancer Inst* **98**, 1528-1537 (2006).
10. Preston, D.L., *et al.* Solid cancer incidence in atomic bomb survivors: 1958-1998. *Radiat Res* **168**, 1-64 (2007).
11. United Nations. 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, Vienna, 2008).
12. Bhatti, P., *et al.* Risk of second primary thyroid cancer after radiotherapy for a childhood cancer in a large cohort study: an update from the childhood cancer survivor study. *Radiat Res* **174**, 741-752 (2010).
13. Rizzo, J.D., *et al.* Solid cancers after allogeneic hematopoietic cell transplantation. *Blood* **113**, 1175-1183 (2009).
14. Travis, L.B., *et al.* Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. *JAMA* **290**, 465-475 (2003).

15. Inskip, P.D., *et al.* Radiation dose and breast cancer risk in the childhood cancer survivor study. *J Clin Oncol* **27**, 3901-3907 (2009).
16. Tucker, M.A., *et al.* Bone sarcomas linked to radiotherapy and chemotherapy in children. *N Engl J Med* **317**, 588-593 (1987).
17. Henderson, T.O., *et al.* Secondary sarcomas in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. *J Natl Cancer Inst* **99**, 300-308 (2007).
18. Lichter, M.D., *et al.* Therapeutic ionizing radiation and the incidence of basal cell carcinoma and squamous cell carcinoma. The New Hampshire Skin Cancer Study Group. *Arch Dermatol* **136**, 1007-1011 (2000).
19. Ron, E., *et al.* Skin tumor risk among atomic-bomb survivors in Japan. *Cancer Causes Control* **9**, 393-401 (1998).
20. Haddy, N., *et al.* Role of radiotherapy and chemotherapy in the risk of secondary leukaemia after a solid tumour in childhood. *Eur. J. Cancer* **42**, 2757-2764 (2006).
21. Inskip, P.D. & Curtis, R.E. New malignancies following childhood cancer in the United States, 1973-2002. *Int J Cancer* **121**, 2233-2240 (2007).
22. Leone, G., Fianchi, L., Pagano, L. & Voso, M.T. Incidence and susceptibility to therapy-related myeloid neoplasms. *Chem Biol Interact* **184**, 39-45 (2010).
23. Travis, L.B., *et al.* Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. *J Natl Cancer Inst* **94**, 182-192 (2002).
24. van den Belt-Dusebout, A.W., *et al.* Roles of radiation dose and chemotherapy in the etiology of stomach cancer as a second malignancy. *Int J Radiat Oncol Biol Phys* **75**, 1420-1429 (2009).
25. Veiga, L.H., *et al.* Chemotherapy and thyroid cancer risk: a report from the Childhood Cancer Survivor Study. *Cancer Epidemiol Biomarkers Prev* (2011).
26. Barnett, G.C., *et al.* Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nat Rev Cancer* **9**, 134-142 (2009).
27. Knight, J.A., *et al.* Genome-wide association study to identify novel loci associated with therapy-related myeloid leukemia susceptibility. *Blood* **113**, 5575-5582 (2009).
28. Varley, J.M. Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat* **21**, 313-320 (2003).
29. Hollander, M.C., Blumenthal, G.M. & Dennis, P.A. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nat Rev Cancer* **11**, 289-301 (2011).
30. Lynch, H.T., *et al.* Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* **76**, 1-18 (2009).
31. Robison, L.L., *et al.* The Childhood Cancer Survivor Study: a National Cancer Institute-supported resource for outcome and intervention research. *J Clin Oncol* **27**, 2308-2318 (2009).
32. Best, T., *et al.* Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. *Nat Med* **17**, 941-943 (2011).
33. Diller, L., *et al.* Chronic disease in the Childhood Cancer Survivor Study cohort: a review of published findings. *J Clin Oncol* **27**, 2339-2355 (2009).
34. Tucker, M.A., *et al.* Leukemia after therapy with alkylating agents for childhood cancer. *J Natl Cancer Inst* **78**, 459-464 (1987).

35. Stovall, M., *et al.* Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies. *Radiat Res* **166**, 141-157 (2006).
36. Lin, D.Y. Cox regression analysis of multivariate failure time data: the marginal approach. *Stat Med* **13**, 2233-2247 (1994).
37. Price, A.L., *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
38. Tukenova, M., *et al.* Radiation therapy and late mortality from second sarcoma, carcinoma, and hematological malignancies after a solid cancer in childhood. *Int J Radiat Oncol Biol Phys* **80**, 339-346 (2011).
39. Sadetzki, S., *et al.* Long-term follow-up for brain tumor development after childhood exposure to ionizing radiation for tinea capitis. *Radiat Res* **163**, 424-432 (2005).
40. Wiemels, J.L., *et al.* Reduced allergy and immunoglobulin E among adults with intracranial meningioma compared to controls. *Int J Cancer* (2011).
41. Papaemmanuil, E., *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1006-1010 (2009).
42. Trevino, L.R., *et al.* Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1001-1005 (2009).

Tables

Table 1: Original and subsequent neoplasm diagnoses among 14,359 children in CCSS.[5] Abbreviations: acute lymphoblastic leukemia (ALL); acute myeloid leukemia (AML); central nervous system (CNS); Hodgkin lymphoma (HL); medulloblastoma/primitive neuroectodermal tumor (medullo/PNET); non-Hodgkin lymphoma (NHL); nonmelanoma skin cancer (NMSC); soft tissue sarcoma (STS).

	Subsequent neoplasm type																	
		Leukemia			Lymphoma			CNS				Solid organ					Skin	
Primary diagnosis	Total	ALL	AML	Other	HL	NHL	Other	Glial	Medullo/PNET	Meningioma	Other	Breast	Bone	STS	Thyroid	Other	Melanoma	NMSC
	Number diagnosed as a second neoplasm																	
ALL	346	2	6		2	3	2	27	2	70	8	12	4	10	15	33	11	139
AML	29	1								1					2	3	3	1
Other leukemia	13							1		3					5	4	4	5
Astrocytoma	85		1	1	1			8		26	1	3	3	5	1	10	3	22
Medullo/PNET	47		1					1		16			1	2	10	4		12
Other CNS	31							2	1	10			1	3	2	3	2	7
HL	453		8	6		10	1	3	1		3	103	6	22	33	56	9	192
NHL	82	2	2		3	1		2		5	3	6	2	3	9	13	2	29
Kidney	51	2										6	4	8	2	12	3	14
Neuroblastoma	45	1	3		1			2				2		4	9	17		6
STS	112			3	2			1		5	1	13	11	17	7	21	4	27
Ewing sarcoma	49	1	2							1		12	9	2	6	7	1	8
Osteosarcoma	55	1		1				2		1		13		3	4	12	4	14
Other bone	4											1				1		2
Total	1402	10	23	11	9	14	3	49	4	135	16	176	41	81	105	197	42	485
	Number diagnosed as a third or higher neoplasm																	
ALL	307									15		3	1	1	6	7	2	272
AML	30									1		1		1		2		25
Other leukemia	10							1								1		8
Astrocytoma	49						1	1		5		1		1		2	1	37
Medullo/PNET	66						1			9								55
Other CNS	52							1		2				1	2	3		43
HL	621						5	1		2		58		4	9	13	4	526
NHL	71													1	1	3		66
Kidney	28								1			1			1			25
Neuroblastoma	4													1	1			2
STS	31		1							1		3	2	1	1	1	4	7
Ewing sarcoma	13											2				1	1	9
Osteosarcoma	19											7	1	2	2	1	2	4
Total	1301	0	1	0	0	0	7	4	1	35	0	76	4	13	23	24	14	1089

Table 2. Estimated availability of biospecimens by primary cancer and specimen type [a]

	Total	Total any biospecimen		Blood		Oragene>1.2ug [b]		Buccal>1.2ug [c]		None>1.2ug	
Primary cancer	N	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
ALL	4329	2127	(49%)	484	(11%)	827	(19%)	612	(14%)	204	(5%)
AML	356	161	(45%)	26	(7%)	63	(18%)	49	(14%)	23	(6%)
Other leukemia	145	42	(29%)	9	(6%)	17	(12%)	13	(9%)	3	(2%)
Astrocytomas	1182	500	(42%)	41	(3%)	232	(20%)	178	(15%)	49	(4%)
Medulloblastoma/PNET	381	182	(48%)	16	(4%)	85	(22%)	57	(15%)	24	(6%)
Other CNS	314	141	(45%)	18	(6%)	60	(19%)	53	(17%)	10	(3%)
HL	1927	953	(49%)	259	(13%)	421	(22%)	111	(6%)	162	(8%)
NHL	1080	514	(48%)	87	(8%)	228	(21%)	168	(16%)	31	(3%)
Kidney	1256	644	(51%)	72	(6%)	311	(25%)	193	(15%)	68	(5%)
Neuroblastoma	955	491	(51%)	66	(7%)	200	(21%)	173	(18%)	52	(5%)
STS	1246	580	(47%)	78	(6%)	264	(21%)	194	(16%)	44	(4%)
Ewings sarcoma	403	188	(47%)	38	(9%)	88	(22%)	46	(11%)	16	(4%)
Osteosarcoma	733	351	(48%)	81	(11%)	147	(20%)	92	(13%)	31	(4%)
Other bone	52	21	(40%)	3	(6%)	8	(15%)	7	(13%)	3	(6%)
Total	14359	6895	(48%)	1278	(9%)	2951	(21%)	1946	(14%)	720	(5%)

[a] Categorization of the specimen type reflects the assumption that DNA quality is likely to be highest in the blood samples, next best in the Oragene samples, and lower in the mouthwash buccal cell samples.

[b] No blood sample available.

[c] No blood or oragene sample available.

Table 3. Estimated availability of biospecimens by primary cancer and type of second neoplasm. Data on number of specimens (N) are derived from January 2012 data files, whereas percentages are based on totals in Friedman et al., 2010,[5] and thus are approximations. Note that analyses plan to consider all subsequent neoplasms rather than only second neoplasms, which will increase sample sizes. Conversely, some available biospecimens may not have sufficient DNA quantity and thus will be excluded from the study.

Primary cancer	Any second neoplasm N (%)[a]	Leukemia N (%)[a]	Lymphoma N (%)[a]	Meningioma N (%)[a]	Other CNS N (%)[a]	Breast N (%)[a]	Bone N (%)[a]	Thyroid N (%)[a]	STS N (%)[a]	NMSC N (%)[a]	Melanoma N (%)[a]	Other Cancers N (%)[a]
ALL	246 (71%)	2 (25%)	4 (57%)	62 (89%)	7 (19%)	9 (75%)	0 -	16 (100%)	5 (50%)	121 (87%)	7 (64%)	13 (39%)
AML	18 (62%)	0 -	0 -	0 -	0 -	4 (100%)	0 -	2 (100%)	2 (100%)	9 (100%)	0 -	2 (67%)
Other leukemia	8 (62%)	0 -	0 -	1 (33%)	0 -	0 -	0 -	5 (100%)	0 -	0 -	0 -	2 (50%)
Astrocytomas	51 (60%)	1 (50%)	1 (100%)	20 (77%)	7 (78%)	1 (33%)	1 (33%)	1 (100%)	2 (40%)	14 (64%)	3 (100%)	0 -
Medullo/PNET	33 (70%)	0 -	0 -	10 (63%)	2 (100%)	0 -	0 -	6 (60%)	2 (100%)	10 (83%)	0 -	3 (75%)
Other CNS	19 (61%)	0 -	0 -	7 (70%)	2 (67%)	0 -	0 -	2 (100%)	1 (33%)	4 (57%)	1 (50%)	2 (67%)
HL	306 (68%)	3 (21%)	6 (55%)	0 -	2 (29%)	78 (76%)	3 (50%)	29 (88%)	10 (45%)	152 (79%)	7 (78%)	16 (29%)
NHL	55 (67%)	1 (25%)	3 (75%)	6 (100%)	0 -	9 (100%)	0 -	5 (56%)	1 (33%)	19 (66%)	1 (50%)	10 (77%)
Kidney	34 (67%)	0 -	0 -	0 -	0 -	1 (17%)	2 (50%)	2 (100%)	4 (50%)	17 (100%)	1 (33%)	7 (58%)
Neuroblastoma	27 (60%)	2 (50%)	2 (100%)	0 -	1 (50%)	1 (50%)	0 -	5 (56%)	1 (25%)	5 (83%)	0 -	10 (59%)
STS	67 (60%)	1 (33%)	2 (100%)	3 (60%)	0 -	10 (77%)	1 (9%)	5 (71%)	10 (59%)	22 (81%)	3 (75%)	10 (48%)
Ewings sarcoma	33 (67%)	1 (50%)	0 -	0 -	0 -	8 (67%)	4 (44%)	4 (67%)	0 -	12 (100%)	2 (100%)	2 (29%)
Osteosarcoma	36 (65%)	1 (50%)	0 -	1 (100%)	1 (50%)	10 (77%)	0 -	3 (75%)	4 (100%)	9 (64%)	1 (25%)	6 (50%)
Other bone	3 (75%)	0 -	0 -	0 -	0 -	1 (100%)	0 -	0 -	0 -	1 (50%)	0 -	1 (100%)
Total	936 (67%)	12 (28%)	18 (69%)	110 (81%)	22 (32%)	132 (75%)	11 (27%)	85 (81%)	42 (52%)	395 (81%)	26 (62%)	84 (43%)

[a] Percentages are based on the total number of patients with a specific primary cancer and second neoplasm, derived from Friedman et al., 2010.[5]

Table 4: Power to detect an association with a SNP that magnifies the effect of radiotherapy on risk of secondary cancer at a genome-wide significance level of 10^{-7}, assuming 50% of patients were exposed to radiation. The odds ratio, comparing the effect of maximum radiation to no radiation, is increased by a factor of 2.0, 2.2, 2.5, 2.7, and 3.0 in individuals with one copy of the minor allele (assuming a high effect size based on Best *et al.*[32]). The risk of cancer was defined by logistic regression where the linear component was β_0+R+ β_GGR, and subsequent neoplasms occurred in 1000 individuals. R is the radiation level, uniformly distributed between 0 and 1 and G is the number of minor alleles.

Increased odds	Case:Control ratio in the study sample (N=1000 cases)				
	1:1	1:2	1:3	1:4	1:5
2.0	0.001	0.015	0.036	0.072	0.086
2.2	0.008	0.065	0.13	0.182	0.221
2.5	0.03	0.15	0.279	0.397	0.447
2.7	0.062	0.328	0.506	0.616	0.702
3.0	0.148	0.534	0.703	0.81	0.858

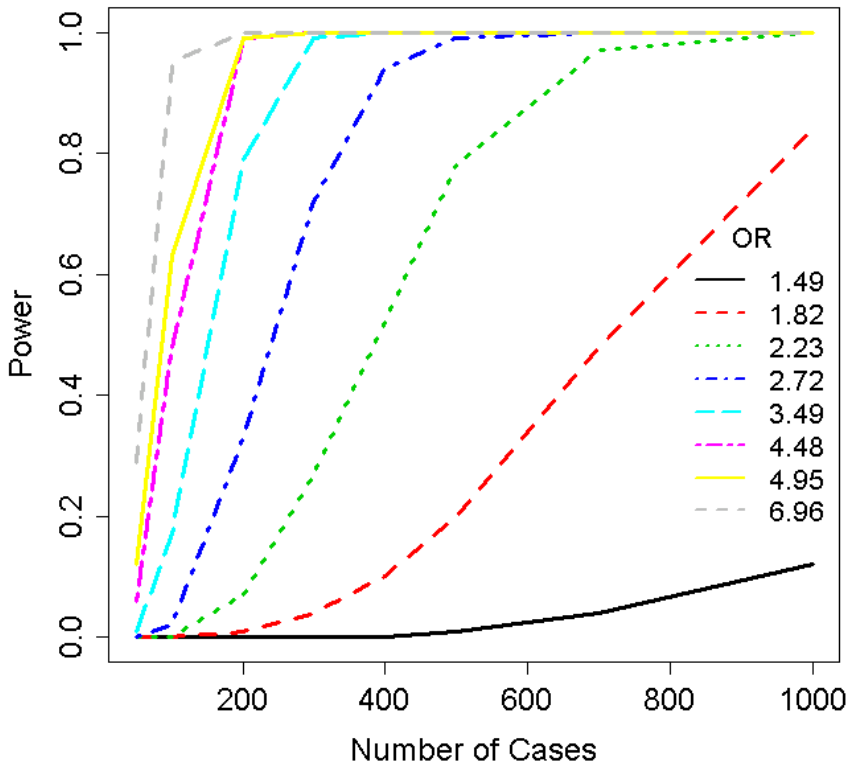
Table 5: Power to detect an association with a SNP that magnifies the effect of radiotherapy on risk of secondary cancer at a genome-wide significance level of 10^{-7}. The odds ratio, comparing the effect of maximum radiation to no radiation, is increased by a factor of 1.5, 1.8, 2.2, 2.7, 3.5, and 4.5 in individuals with one copy of the minor allele. The risk of cancer was defined by logistic regression where the linear component was β_0+R+ β_GGR, and secondary cancers occurred in 1000 individuals. R indicates whether the subject received radiation (e.g. subjects only received high or no dose) and G is the number of minor alleles. We assume that 70% of the subjects received radiation.

Increased odds	Case:Control ratio in the study sample (N=1000 cases)				
	1:1	1:2	1:3	1:4	1:5
1.5	0.01	0.044	0.0797	0.1089	0.1258
1.8	0.2596	0.6112	0.7617	0.836	0.8673
2.2	0.8622	0.9899	0.9976	0.9995	0.9996
2.7	0.9983	1	1	1	1
3.5	1	1	1	1	1
4.5	1	1	1	1	1

Table 6: Power to detect an association with a SNP that magnifies the effect of radiotherapy on risk of secondary cancer at a genome-wide significance level of 10^{-7}. The odds ratio, comparing the effect of maximum radiation to no radiation, is increased by a factor of 1.5, 1.8, 2.2, 2.7, 3.5, and 4.5 in individuals with one copy of the minor allele. The risk of cancer was defined by logistic regression where the linear component was β_0+R+ β_GGR, and secondary cancers occurred in 1000 individuals. R indicates the magnitude of radiation and is uniformly distributed between 0 and 1, and G is the number of minor alleles. We assume that 70% of the subjects received radiation and that the ***SNP only influences risk for 50% of the possible second neoplasms.***

Increased odds	Case:Control ratio in the study sample (N=1000 cases)				
	1:1	1:2	1:3	1:4	1:5
1.5	0	0	0	1.00E-04	1.00E-04
1.8	1.00E-04	0	5.00E-04	4.00E-04	5.00E-04
2.2	4.00E-04	0.0024	0.0056	0.0078	0.0096
2.7	0.0031	0.0175	0.0413	0.0582	0.078
3.5	0.0313	0.1432	0.2693	0.3598	0.4128
4.5	0.1538	0.4982	0.7044	0.802	0.8535

Figure 1. If we assume that the case:control ratio is 1:5, then power is only a function of the number of cases meeting the stated criteria (e.g., a specific primary cancer and specific type of subsequent neoplasm, or number of patients treated with radiotherapy to a particular region of the body). In the figure below, we illustrate power as a function of number of cases for different effect sizes, with MAF fixed at 0.1 and assuming that the criteria for defining the case subset was specified *a priori*. All other parameters are the same as those described for Table 5.



Appendix: Power Calculations

To obtain an estimate for the power to detect an association between SNP and subsequent neoplasm, we assumed that radiation affected all cancers equally and its effect increased by a factor of β_1 on the logistic scale for individuals with at least one minor allele. The employed model was

$$logit(\mu) = \beta_0 + X + \beta_1 GX$$

where X is radiation dose, G indicates presence of minor allele, and β_0 is chosen so the overall rate of subsequent neoplasms is 0.1. We assumed that 50% of the 10,000 individuals in the cohort received no radiation and the remaining doses were uniformly distributed between 0 and 1. The association between SNP and rate of subsequent neoplasms was tested by comparing logistic-regression models with and without the gene and gene-radiation interaction terms. Power was estimated by simulation at an α-level = 10^{-7}.