1. STUDY TITLE

Germline genetic variation and the risk of chemotherapy-associated subsequent malignant neoplasms for survivors of childhood cancers

2. WORKING GROUPS AND INVESTIGATORS

Genetics Working Group

Smita Bhatia, MD, MPH (smitabhatia@uabmc.edu)

Second Malignancy Working Group

Joseph P. Neglia, MD, MPH (jneglia@umn.edu)
Lucie M. Turcotte, MD, MPH (turc0023@umn.edu)

Memorial Sloan Kettering Cancer Center (host institution)

Gordon P. Watt, PhD (wattg@mskcc.org)
Jonine L. Bernstein, PhD (bernstej@mskcc.org)
Chaya S. Moskowitz, PhD (moskowc1@mskcc.org)

Other CCSS Investigators
Greg T. Armstrong, MD (Greg.Armstrong@stjude.org)
Yutaka Yasui, PhD (Yutaka.Yasui@stjude.org)
Wendy M. Leisenring, ScD (wleisenr@fredhutch.org)
Michael Arnold, MD (Michael.arnold@childrenscolorado.org)
Miriam Conces, MD (Miriam.conces@nationwidechildrens.org)

3. BACKGROUND AND RATIONALE

Background. Chemotherapeutic drugs for childhood cancers improve survival, but also increase the risk of subsequent malignant neoplasms (SMNs).[1, 2] The Childhood Cancer Survivor Study (CCSS), St. Jude Lifetime Cohort Study (SJLIFE), and others have reported associations of anthracyclines generally, and doxorubicin dose specifically, with subsequent breast cancer;[2-4] cyclophosphamide equivalent dose, ifosfamide, and daunorubicin with subsequent sarcoma;[1, 4] cumulative anthracycline dose with thyroid cancer;[4] and procarbazine and platinum with gastrointestinal (GI) SMNs.[5] Of note, reductions in radiation dose/volume used for treating childhood cancers in recent decades[6] coincide with an increased intensity of some chemotherapy regimens[7]. Accordingly, there is a need to identify individuals with high risk of chemotherapy-related SMNs to inform long-term surveillance.

Limited genetic risk factors for chemotherapy-related SMNs have been identified. Individuals with cancer predisposition syndromes[8] or other high-penetrance mutations have an increased risk of SMNs with or without chemotherapy,[9] though the population risk of SMNs attributable to these rare mutations is small. At least 20% of the remaining variation in risk may be explained by lower-penetrance mutations that together confer an increased risk of chemotherapy-associated SMNs.[10] For example, polygenic risk scores (PRSs) for subsequent breast and thyroid cancers have improved risk prediction for these SMNs.[11, 12] These PRSs comprised SNPs that were associated with risk of *de novo* breast/thyroid cancers, and may not fully capture the distinct etiology of chemotherapy-associated SMNs. Candidate gene studies have reported associations of chemotherapy-associated SMNs with missense mutations in drug metabolism genes,[13] DNA repair genes,[14, 15] and telomere maintenance genes,[16] Further study is needed to enable SMN risk prediction for survivors of childhood cancers who received chemotherapy. We address this gap by proposing to identify the genetic risks of SMNs for childhood cancer survivors who received chemotherapy but no radiation (hereafter, “chemotherapy-alone”) for the treatment of childhood cancer.

The purpose of this study is to (A) determine whether exonic genetic variants are associated with SMN risk for survivors who received chemotherapy-alone for their childhood cancer and (B) identify exonic genetic variants that interact with cumulative chemotherapy dose to increase the risk of SMNs.

4. SPECIFIC AIMS

Specific Aim 1. Identify exonic genetic variants associated with chemotherapy-associated SNMs for childhood cancer survivors. ***Aim 1.1. Discovery study in the CCSS.*** We will test the association of solid SMN risk with exonic genetic variation for participants in the Original and Expansion CCSS Studies who received chemotherapy but not radiotherapy for their childhood cancer. In exploratory analyses, the associations will also be evaluated separately for two groups defined by class of chemotherapy drugs received: the anthracycline group and the alkylating agent group (see Analysis Framework, section 5D). Finally, exploratory subset analysis will be completed to identify genetic variants associated with specific chemotherapy-associated SMNs: (a) breast cancers; sarcomas; and thyroid cancers. ***Aim 1.2. Replication study in the SJLIFE Study.*** We will validate genome-wide significant associations in the St. Jude Lifetime (SJLIFE) Study. For replicated associations, we will investigate the potential mechanisms underlying the association using bioinformatic analysis to identify potentially deleterious effects on protein structure, effects on chromatin organization, or changes in gene expression, among other effects.

Specific Aim 2. Evaluate interactions between chemotherapy dose and genetic variants that increase the risk of SMNs. For each variant with a potential association ($p<1\times10^{-7}$) with SMN risk, we will test the interaction of the variant with chemotherapy dose in multivariable models of SMN risk. The interactions will be evaluated in the overall population and in exploratory subgroups defined by class of chemotherapy drugs received (anthracyclines or alkylating agents); and type of SMN, as defined in Aim 1. For each significant interaction, we will evaluate the nature of the interaction by determining SMN risk for each combination of variant allele and chemotherapy dose. Finally, significant interactions will be validated in the SJLIFE Study.

Specific Aim 3. Develop and test a polygenic risk score (PRS) for chemotherapy-associated SMN risk. For each variant with a potential association with SMN risk ($p<1\times10^{-7}$) identified using the CCSS data in Aim 1, we will construct a PRS for SMN risk by combining the effect estimates of each variant into a single score. The primary PRS will be constructed for the overall associations; exploratory PRSs will be developed for each exposure subgroup (anthracyclines or alkylating agents); and for each specific SMN subgroup. Next, using the SJLIFE Study, we will determine whether the PRSs improve the discrimination of SMN risk beyond known risk factors, including age at diagnosis, type of childhood cancer, and types and dose of chemotherapies received.

Expected outcomes. We expect to understand whether genetic variation is associated with or modifies the risk of chemotherapy-associated SMNs. The long-term goals of this research are to (1) estimate the risk of SMNs at the time of diagnosis to understand risks and benefits of treatment and (2) personalize surveillance for SMNs after receiving chemotherapy for a childhood cancer.

5. ANALYSIS FRAMEWORK

Table 1, below, describes the overall study populations based on inclusion criteria: receipt of chemotherapy-only for treatment and availability of genomic data for analysis. All counts are at the participant level. The study population for the primary analysis is highlighted in yellow.

Table 1. Number of participants and SMNs diagnosed in the study population

	CCSS[a] (Discovery)	SJLIFE (Validation)
Primary analysis		
(A) Participants receiving chemotherapy-only[b]	8115	2360
(B) Chemotherapy-only participants with genomic data[c]	2521	1814
(C) Number of participants in (B) with ≥1 SMN	126	108
Subgroups of (C) for exploratory analyses		
(D) *SMN Subgroups*		
Subsequent breast neoplasm	34	29
Subsequent thyroid neoplasm	25	21
(E) *Chemotherapy categories subgroups*[d]		
(1) Patients who received anthracyclines	1367	1363
Number of patients with ≥1 SMN	82	79
Breast	27	26
Thyroid	18	15
(2) Patients who received alkylating agents	1407	1304
Number of patients with ≥1 SMN	84	86

[a] Includes participants in the Original and Expansion CCSS cohorts.
[b] All treatment exposures are within the 5 years following childhood cancer diagnosis.
[c] Includes participants with either whole-exome or whole-genome sequencing, based on data available in the St Jude Cloud Genomic Browser.
[d] Groups are not mutually exclusive

5A. Discovery study population: the CCSS. The source population includes all participants in the Original and Expansion CCSS cohort Studies (diagnosed between 1970 and 1999) who (a) received chemotherapy but not radiation therapy for their childhood cancer; (b) have cumulative chemotherapy dose for the first five years after diagnosis of the childhood cancer available; and (c) have genomic data (whole genome or whole exome sequencing) available for analysis (n = 2,521). Exploratory subgroup analyses will be completed for the *anthracycline group* (n = 1,367) and the *alkylating agents group* (n = 1,407).

5B. Replication study population: the SJLIFE Study. The source population for the replication study includes all participants in the SJLIFE Study: childhood cancer survivors treated at St Jude Children's Research Hospital between 1962 and 2012 who received chemotherapy but not radiation therapy (n=2,360). Among these chemotherapy-only patients, 1,814 participants have genomic data available for analysis. Participants dual-enrolled in CCSS and SJLIFE will be excluded from the replication study.

5C. Genetic variants. We will evaluate exonic germline variants measured using whole-exome or whole-genome sequencing. We will include single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations (hereafter, "variants"). Any variant whose minor allele is present in less than 2 participants will be excluded. Variants will be evaluated as predictors of SMN risk (Aim 1) and as effect modifiers of the association between chemotherapy dose and SMN risk (Aim 2).

5D. Treatment exposure: chemotherapy dose. In Aim 2, we will assess the statistical interaction between genetic variants and cumulative chemotherapy dose. Because chemotherapeutic agents have specific associations with each SMN site, exploratory analyses will be done in two non-exclusive chemotherapy groups: the anthracycline group will examine the association between anthracycline dose (doxorubicin isotoxic equivalent dose) and risk of breast or thyroid cancer or sarcoma; the alkylating agent group will examine the association of alkylating agent dose (cyclophosphamide equivalent dose[17]) with sarcoma, GI cancers or other carcinomas.

5E. Outcome: Non-hematological SMNs. The primary outcome of interest is diagnosis with any non-hematological ("solid") SMN excluding non-melanoma skin cancer >5 years after diagnosis of the childhood cancer. Within the study population (chemotherapy-only with genomic data available), there are 126 patients diagnosed with at least one SMN in the CCSS Study and 108 in the SJFLIFE Study. The most common SMNs in the CCSS were breast carcinoma, thyroid carcinoma, and melanoma. In the primary analyses, we will group all solid SMNs together. In the exploratory analysis by chemotherapy subgroups analysis, we will group

participants based on the class of chemotherapy drugs received, as described above. Finally, in the exploratory outcomes subgroup analysis, we will evaluate the risk of breast cancers and thyroid cancers separately.

5F. Covariates, confounders, and effect modifiers. In all multivariable analyses, we will adjust for age at time of diagnosis of the first cancer, study recruitment phase, and the top 3 principal components to account for potential population stratification. We will evaluate the presence of confounding and effect modification for the following variables:

a. Type of chemotherapy. For both Aims, we will conduct subgroup analysis for the two main chemotherapy groups: anthracyclines and alkylating agents. We have defined specific SMNs associated with each chemotherapy group (see 5D).

b. Type of childhood cancer. A prior study in the CCSS[2] showed that the risk of SMNs for survivors treated with chemotherapy-only differs by the type of childhood cancers: survivors of leukemias and central nervous system tumors have the greatest risk, followed by survivors of sarcomas, while survivors of Wilms tumor, for instance, have a lower risk. In this study, we will explore whether any genetic associations with SMNs are modified by the type of childhood cancer.

c. Sex. Female sex is associated with SMNs for CCSS participants treated with chemotherapy only. The most common solid SMN is breast cancer, which is diagnosed almost exclusively in female patients. It is not clear whether sex may modify the genetic associations with non-breast SMNs or the interaction with chemotherapy dose.

5E. Approach

Specific Aim 1. Identify exonic genetic variants associated with chemotherapy-associated SNMs for childhood cancer survivors

Aim 1.1. Discovery study in the CCSS. The CCSS genomic data have been prepared and aligned with the reference human genome. We will quantify population substructure within the chemotherapy-only study population by calculating principal components. The top three principal components will be included as covariates in the multivariable models.

Statistical analysis. In the primary analysis, all participants who received chemotherapy-only will be included and the outcome will the diagnosis of any solid SMN. First, we will test for independence of each variant with SMN risk using the Mantel-Haenszel (MH) test statistic. The MH test statistic is not subject to asymptotic assumptions that are unlikely to be satisfied within this subset of participants receiving chemotherapy-only. The MH test will be based on a 2×*n*×2 table with 2 strata defined by cohort (Original or Expansion); within each stratum, the *n*×*2* tables represent the variant alleles (*n*≥2) and SMN status (yes or no). For variants associated with SMN diagnosis with $P_{MH}<1x10^{-6}$, we will further evaluate the association between each variant and SMN risk in a multivariable Cox proportional hazards model assuming additive genetic effects. Age will be used as the time scale, with each participant followed until the first SMN diagnosis, with censoring at the time of last follow-up, death, or diagnosis of an SMN that is not an outcome of interest (e.g., leukemia). Adjustment variables will include age at time of diagnosis of the childhood cancer, cohort (Original or Expansion), and three principal components for population stratification. We will report the hazards ratio (HR) and 95% confidence interval (CI) for each of the "top" associations ($P_{MH}<1\times10^{-7}$). Finally, we will calculate permutation-based *P*-values for the multivariable associations as done in Morton et al. *JNCI* (2017).[18] Variants with permutation-based $P<5\times10^{-8}$ will be considered significant at the genome-wide level and will be validated in the SJLIFE Study.

Table 2. Power for discovery analysis

Allele frequency	Θ_{MH}	Power
0.2	2.00	0.18
	2.25	0.64
	2.75	*0.80*
0.1	2.00	0.03
	3.00	0.61
	3.32	*0.80*
0.05	3.00	0.10
	4.00	0.57
	4.53	*0.80*

Power calculation. Assuming an additive genetic model, we present the estimated power (Table 2) to reject the null hypothesis in the MH tests with $P_{MH}<1x10^{-6}$ for distinct allele frequencies and MH Θ statistic (analogous to odds ratio). For common variants (MAF ≥ 0.2), we achieve 80% power at an effect size of Θ = 2.75; for rare variants (MAF < 0.05), we achieve 80% power at Θ = 4.53. In a multiplicative model, or for allele frequencies >0.2, 80% power is achieved at lower effect sizes. In a prior genome-wide association study of radiation-associated subsequent breast neoplasms in the CCSS, the top two SNP associations in the irradiated group had HRs of 1.92 and 2.47 (MAF in controls 0.47 and 0.02, respectively), suggesting that this study is powered to detect realistic associations between genetic variants and SMNs in the

chemotherapy-only subgroup.

Exploratory analysis. As described in 5E, we will test for genetic associations separately for the anthracycline group (outcomes: breast carcinoma, thyroid carcinoma, sarcoma) and the alkylating agent group (gastrointestinal cancers, sarcomas, and other carcinomas). These subgroups are not mutually exclusive. Further, we will statistically test for differences in the top genetic associations by type of childhood cancer (hematologic vs non-hematologic) and sex. Finally, the analyses will be repeated in two exploratory SMN subgroups: breast cancers (limited to female participants) and thyroid cancers.

Aim 1.2. Replication study in the SJLIFE Study. We will conduct replication analysis to confirm genome-wide significant variants (CCSS permutation-based *p*-value $<5\times10^{8}$) in the SJLIFE Study. Using Cox proportional hazards models, we will assess the association between SMNs and genetic variants, adjusting for age at childhood cancer diagnosis, and three principal components to account for population stratification, with censoring and follow-up as described above.

For any variants that are validated in the SJLIFE Study, we will perform bioinformatic analysis to identify and describe potential mechanisms of increased SMN risk, including effects on protein structure, alternative splicing, histone modification, and transcriptional factor binding. We will also determine whether the identified variants are in linkage disequilibrium with non-coding regions, which may in turn impact cis-regulatory elements and other transcriptional enhancers.

Specific Aim 2. Evaluate interactions between chemotherapy dose and genetic variants that increase the risk of SMNs.

The objective of Aim 2 is to evaluate statistical interactions between the statistically significant variants identified in Aim 1 and cumulative chemotherapy dose received for the childhood cancer in the association with SMN risk.

Statistical approach. The therapies in the anthracycline group have been previously converted to doxorubicin isotoxic equivalent dose (mg/m^2) and the therapies in the alkylating agent group have been converted to cyclophosphamide equivalent dose (mg/m^2). The dosages for anthracyclines and alkylating agents are not directly comparable, with much higher absolute doses for alkylating agents. Further, the mechanisms of action and the associations with SMNs differ between the groups. Therefore, interaction analyses will be completed separately for participants who received anthracyclines and participants who received alkylating agents. A small proportion (<5%) of participants received both chemotherapies and will be included in both groups.

For each variant, we will develop a Cox proportional hazards model, as in Aim 1, to assess the interaction with cumulative chemotherapy dose with adjustment for age at time of diagnosis of the childhood cancer, CCSS Study phase (Original or Expansion), and the top three principal components. *We will interpret only the interaction effect in the multivariable Cox model*. We will determine statistical significance via likelihood ratio test and apply the Benjamini-Hochberg procedure to reduce the false discovery rate to <10%. Statistically significant interactions will be validated in the SJLIFE Study. Validated interactions will be further interrogated to understand the direction and nature of the interaction with chemotherapy. Finally, we will use bioinformatic tools, as in Aim 1, to develop hypotheses about the potential mechanisms underlying the interactions.

Specific Aim 3. Develop and test a PRS for chemotherapy-associated SMN risk.

The objective of Aim 3 is to combine the individual effects of variants potentially associated with SMN risk to create a PRS that improves the discrimination of SMN risk beyond known risk factors.

Statistical approach. Using the variants that had an association with SMN with a suggestive permutation-based p-value$<5\times10^{-6}$ in the CCSS in Aim 1, we will construct a PRS. The PRS will be constructed as a linear combination of each variant (g_i) multiplied by the log of the HR from the multivariable-adjusted Cox proportional hazards model, $PRS = g_1\times\log(HR_1) + \ldots + g_i\times\log(HR_i)$. The final score will then be used to assign a PRS to each participant in the SJLIFE Study population. It is expected that the PRS will have an approximately Normal distribution but may be scaled and/or transformed prior to analysis. To assess the relative improvement in discrimination of SMN risk using the PRS, we will calculate multivariable area under the receiver operating characteristic (AUROC) for models with and without the PRS. To statistically test the difference in AUROC, we will use the approach developed by Seshan et al.[19]

6. APPENDIX.

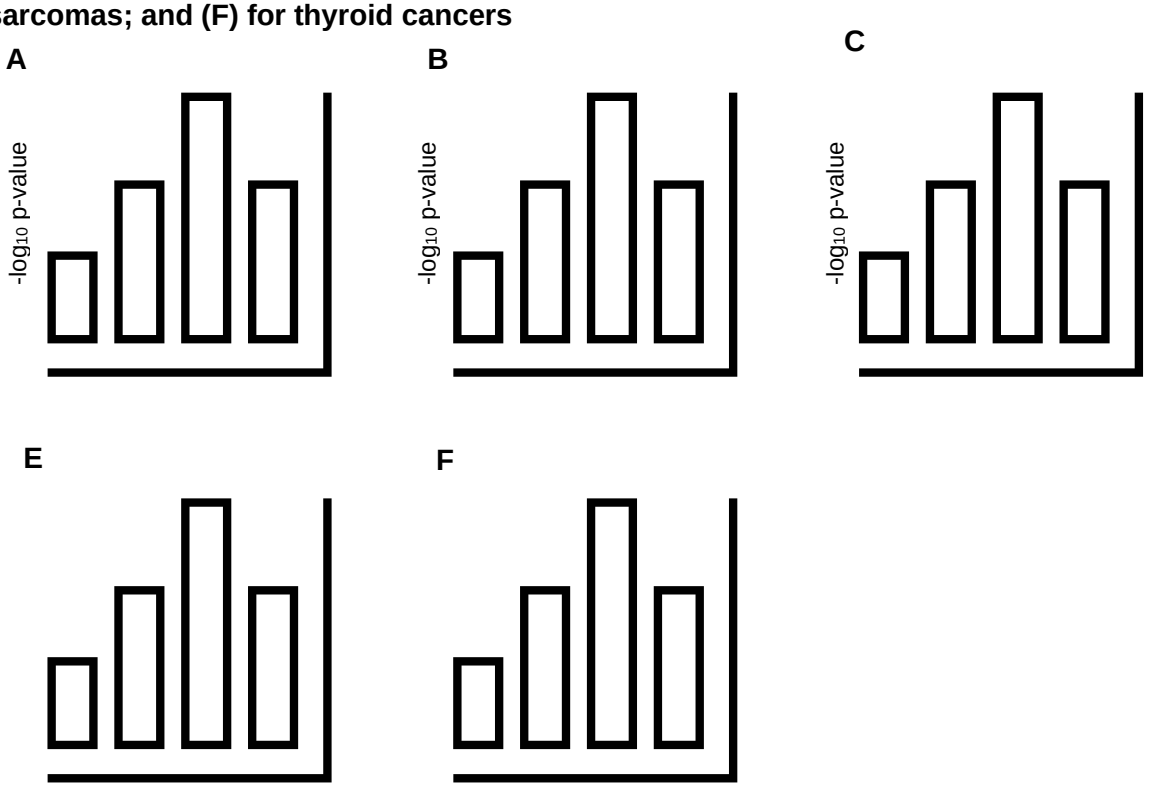
Potential tables and figures for Aim 1.

Aim 1, Table 1. Description of study populations in the CCSS and SJLIFE Studies who received chemotherapy-only for their childhood cancer and have genomic data available for analysis.

Characteristic[a]	CCSS, (N = 2,521)	SJLIFE, (N = 1,814)
Sex		
Male		
Female		
Race/ethnicity		
Non-Hispanic white		
Non-Hispanic Black		
Non-Hispanic Asian		
Hispanic (any race)		
Other		
Age (y) at childhood cancer diagnosis		
0-4		
5-9		
10-14		
≥15		
Decade of treatment		
1960-1969		
1970-1979		
1980-1989		
1990-1999		
2000-2012		
Total follow-up time, years		
5-9		
10-14		
14-19		
≥20		
Childhood cancer diagnosis		
Leukemia		
CNS tumor		
Hodgkin Lymphoma		
Non-Hodgkin Lymphoma		
Neuroblastoma		
Soft-tissue sarcoma		
Bone cancer		
Chemotherapy class received		
Anthracycline(s) only		
Alkylating agent(s) only		
Anthracyclines and alkylating agent		
Type of sequencing and biospecimen type		
Whole exome		
Buccal cells		
Whole blood		
Saliva		
Whole genome		
Buccal		
Whole blood		
Saliva		

Abbreviations. CCSS, Childhood Cancer Survivor Study; SJLIFE, St. Jude Lifetime Cohort Study; CNS, central nervous system

Aim 1, Figure 1. Manhattan plots of associations between exonic variants and SMN risk, (A) overall; (B) for the anthracycline group; (C) for the alkylating agents group; (D) for breast cancers; (E) for sarcomas; and (F) for thyroid cancers



Note. The anthracycline group includes participants who received anthracyclines-only (outcomes: breast cancer, thyroid cancer, or sarcoma) and the alkylating agents group includes participants who received only alkylating agents (outcomes: GI cancers, sarcomas, other carcinomas) within 5 years of the childhood cancer diagnosis.

Aim 1, Table 2. Details of the variants with $p < 5x10^{-7}$ in multivariable Cox proportional hazards model of SMN risk

Variant	Gene	Type of variant[a]	Chr:pos	Minor allele frequency (%)	Hazard ratio (95% CI)[b]	Multivariable p-value[c]	Mantel-Haenszel p-value[c]
Variant 1							
…							
Variant n							

[a] Single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations
[b] Per-allele effect from Cox proportional hazards model adjusting for age at diagnosis of the childhood cancer, study phase (Original CCSS or Expansion Study), and top three principal components
[c] Permutation-based *p*-value
[c] From two-sided Mantel-Haenszel test from exact conditional distribution, stratified on CCSS study phase.

Aim 1, Table 3. Replication study of variants in the SJLIFE Study

Variant	Gene	Type of variant[a]	Chr:pos	Minor allele frequency (%)	Hazard ratio (95% CI)[b]	Multivariable p-value[c]
Variant 1						
…						
Variant n						

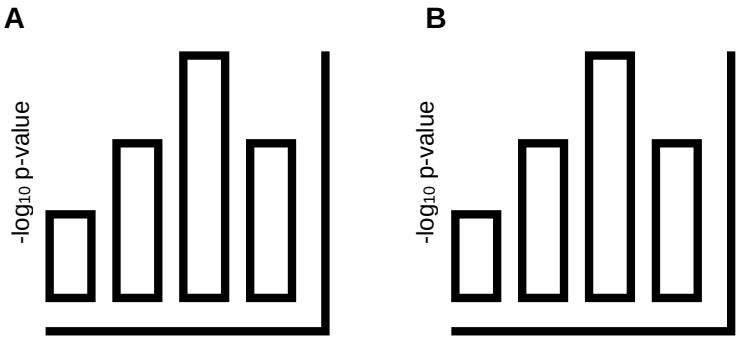
[a] Single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations
[b] From Cox proportional hazards model adjusting for age at diagnosis of the childhood cancer, study phase (original or expansion), and top three principal components
[c] Permutation-based *p*-value

Aim 1, Figure 2. Regional mapping of the replicated variants associated with SMN risk for participants who received chemotherapy-only

[LocusZoom plot or similar]

Potential Tables and Figures for Aim 2.

Aim 2, Figure 1. Manhattan plots of the genotype × chemotherapy dose interactions associated for SMN risk (A) for the anthracycline group and (B) the alkylating agent group in the CCSS Study.



Note. The anthracycline group includes participants who received anthracyclines-only (outcomes: breast cancer, thyroid cancer, or sarcoma) and the alkylating agents group includes participants who received only alkylating agents (outcomes: GI cancers, sarcomas, other carcinomas) within 5 years of the childhood cancer diagnosis.

Aim 2, Table 1. Statistically significant genotype × chemotherapy dose interaction effects in the CCSS Study

Variant	Gene	Type of variant[a]	Chr:pos	Minor allele frequency (%)	Hazard ratio (95% CI)[b]	Multivariable p-value[b]	Interpretation
Anthracycline group							
Variant 1							
…							
Variant n							
Alkylating agent group							
Variant 1							
…							
Variant n							

[a] single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations
[b] Coefficient for interaction of genetic variant with chemotherapy dose (in mg/m^2) from Cox proportional hazards model adjusting for age at diagnosis of the childhood cancer, study phase (Original CCSS or Expansion Stuy), and top three principal components

Aim 2, Table 2. Replication study of interactions in the SJLIFE Study

Variant	Gene	Type of variant[a]	Chr:pos	Minor allele frequency (%)	Hazard ratio (95% CI)[b]
Variant 1					
…					
Variant n					

[a] single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations
[b] Coefficient for interaction of genetic variant with chemotherapy dose (in mg/m^2) from Cox proportional hazards model adjusting for age at diagnosis of the childhood cancer, study phase (original or expansion), and top three principal components

Aim 2, Table 3. Details of effect modification by chemotherapy dose for replicated interactions

Variants	Chemotherapy dose < median		Chemotherapy dose ≥ median		P-value for heterogeneity[b]
	HR (95% CI)[a]	P-value[a]	HR (95% CI)[a]	P-value[a]	
Variant 1					
…					
Variant n					

[a] Hazard ratio (HR) and P-value for the interaction between each variant and dichotomized chemotherapy dose in multivariable Cox proportional hazards model
[b] P-value from likelihood ratio test comparing models with and without interactions between genetic variant and dichotomized chemotherapy dose

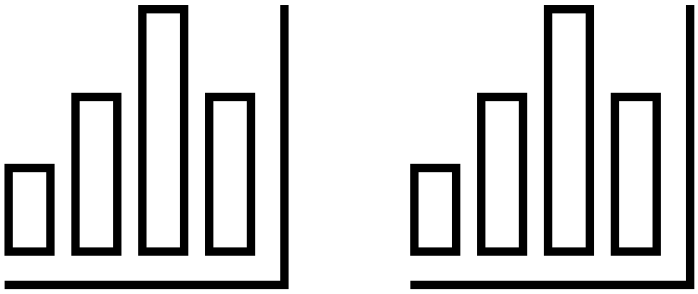
Potential Tables and Figures for Aim 3.

Aim 3, Table 1. Description of variants included in PRS

Variant	Gene	Type of variant[a]	Chr:pos	Minor allele frequency (%)	Log(HR)[b]
Variant 1					
…					
Variant n					

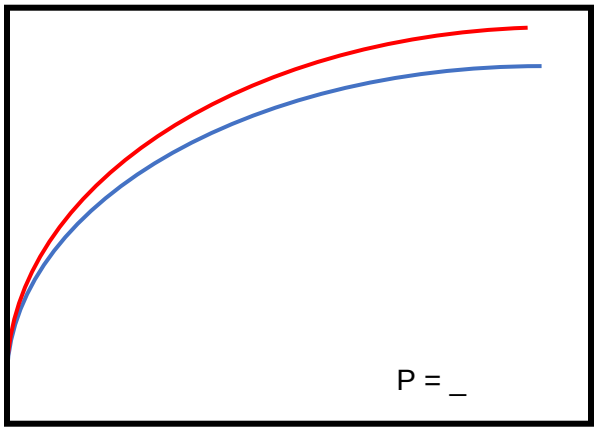
[a] single nucleotide polymorphisms, small insertions and deletions, multi-allelic variants, and copy number variations
[b] Coefficient for association between (in mg/m^2) from Cox proportional hazards model adjusting for age at diagnosis of the childhood cancer, study phase (original or expansion), and top three principal components

Aim 3, Figure 1. Distribution (histograms) of PRS in the CCSS and SJLIFE studies



CCSS SJLIFE

Aim 3, Figure 2. AUROC curves for SMN risk in the SJLIFE Study with and without the PRS



Potential supplementary figures and tables

Supplementary Table 1. Comparison of chemotherapy-only study population with and without genomic data available in the CCSS and SJLIFE Studies

Characteristic[a]	CCSS		SJLIFE	
	With genomic data (N = 2,521)	Without genomic data (N = 5,594)	With genomic data (N = 1,814)	Without genomic data (N = 546)
Sex				
Male				
Female				
Race/ethnicity				
Non-Hispanic white				
Non-Hispanic Black				
Non-Hispanic Asian				
Hispanic (any race)				
Other				
Age (y) at childhood cancer diagnosis				
0-4				
5-9				
10-14				
≥15				
Decade of treatment				
1960-1969				
1970-1979				
1980-1989				
1990-1999				
2000-				
Total follow-up time, years				
5-9				
10-14				
14-19				
≥20				
Childhood cancer diagnosis				
Leukemia				
CNS tumor				
Hodgkin Lymphoma				
Non-Hodgkin Lymphoma				
Neuroblastoma				
Soft-tissue sarcoma				
Bone cancer				
Chemotherapy class received				
Anthracycline(s) only				
Alkylating agent(s) only				
Anthracycline(s) and alkylating agent(s)				
Type of sequencing and biospecimen type				
Whole exome				
Buccal cells				
Whole blood				
Saliva				
Whole genome				
Buccal				
Whole blood				
Saliva				

Abbreviations. CCSS, Childhood Cancer Survivor Study; SJLIFE, St. Jude Lifetime Cohort Study.

7. **REFERENCES**

1. Henderson TO, Rajaraman P, Stovall M, Constine LS, Olive A, Smith SA, Mertens A, Meadows A, Neglia JP, Hammond S, Whitton J, Inskip PD, Robison LL, Diller L. Risk factors associated with secondary sarcomas in childhood cancer survivors: a report from the childhood cancer survivor study. Int J Radiat Oncol Biol Phys. 2012;84(1):224-30. Epub 2012/07/17. doi: 10.1016/j.ijrobp.2011.11.022. PubMed PMID: 22795729; PMCID: PMC3423483.
2. Turcotte LM, Liu Q, Yasui Y, Henderson TO, Gibson TM, Leisenring W, Arnold MA, Howell RM, Green DM, Armstrong GT, Robison LL, Neglia JP. Chemotherapy and Risk of Subsequent Malignant Neoplasms in the Childhood Cancer Survivor Study Cohort. J Clin Oncol. 2019;37(34):3310-9. Epub 2019/10/18. doi: 10.1200/jco.19.00129. PubMed PMID: 31622130; PMCID: PMC7001784.
3. Ehrhardt MJ, Howell CR, Hale K, Baassiri MJ, Rodriguez C, Wilson CL, Joshi SS, Lemond TC, Shope S, Howell RM, Wang Z, Srivastava D, Mulrooney DA, Zhang J, Robison LL, Ness KK, Hudson MM. Subsequent Breast Cancer in Female Childhood Cancer Survivors in the St Jude Lifetime Cohort Study (SJLIFE). J Clin Oncol. 2019;37(19):1647-56. doi: 10.1200/jco.18.01099. PubMed PMID: 31075046.
4. Teepen JC, Leeuwen FEv, Tissing WJ, Broeder EvD-d, Heuvel-Eibrink MMvd, Pal HJvd, Loonen JJ, Bresters D, Versluys B, Neggers SJCMM, Jaspers MWM, Hauptmann M, Loo MvdH-vd, Visser O, Kremer LCM, Ronckers CM, Group tDLS. Long-Term Risk of Subsequent Malignant Neoplasms After Treatment of Childhood Cancer in the DCOG LATER Study Cohort: Role of Chemotherapy. J Clin Oncol. 2017;35(20):2288-98. doi: 10.1200/jco.2016.71.6902. PubMed PMID: 28530852.
5. Henderson TO, Oeffinger KC, Whitton J, Leisenring W, Neglia J, Meadows A, Crotty C, Rubin DT, Diller L, Inskip P, Smith SA, Stovall M, Constine LS, Hammond S, Armstrong GT, Robison LL, Nathan PC. Secondary gastrointestinal cancer in childhood cancer survivors: a cohort study. Annals of internal medicine. 2012;156(11):757-260. doi: 10.7326/0003-4819-156-11-201206050-00002. PubMed PMID: 22665813.
6. Turcotte LM, Liu Q, Yasui Y, Arnold MA, Hammond S, Howell RM, Smith SA, Weathers RE, Henderson TO, Gibson TM, Leisenring W, Armstrong GT, Robison LL, Neglia JP. Temporal Trends in Treatment and Subsequent Neoplasm Risk Among 5-Year Survivors of Childhood Cancer, 1970-2015. JAMA. 2017;317(8):814-24. Epub 2017/03/01. doi: 10.1001/jama.2017.0693. PubMed PMID: 28245323; PMCID: PMC5473951.
7. Hudson MM, Neglia JP, Woods WG, Sandlund JT, Pui CH, Kun LE, Robison LL, Green DM. Lessons from the past: opportunities to improve childhood cancer survivor care through outcomes investigations of historical therapeutic approaches for pediatric hematological malignancies. Pediatr Blood Cancer. 2012;58(3):334-43. Epub 2011/11/01. doi: 10.1002/pbc.23385. PubMed PMID: 22038641; PMCID: PMC3256299.
8. Talwalkar SS, Yin CC, Naeem RC, Hicks MJ, Strong LC, Abruzzo LV. Myelodysplastic syndromes arising in patients with germline TP53 mutation and Li-Fraumeni syndrome. Arch Pathol Lab Med. 2010;134(7):1010-5. Epub 2010/07/01. doi: 10.5858/2009-0015-oa.1. PubMed PMID: 20586629.
9. Waespe N, Belle FN, Redmond S, Schindera C, Spycher BD, Rössler J, Ansari M, Kuehni CE. Cancer predisposition syndromes as a risk factor for early second primary neoplasms after childhood cancer - A national cohort study. Eur J Cancer. 2021;145:71-80. Epub 2021/01/11. doi: 10.1016/j.ejca.2020.11.042. PubMed PMID: 33423008.
10. Bhatia S. Genetic variation as a modifier of association between therapeutic exposure and subsequent malignant neoplasms in cancer survivors. Cancer. 2015;121(5):648-63. Epub 2014/10/31. doi: 10.1002/cncr.29096. PubMed PMID: 25355167; PMCID: PMC4339370.
11. Song N, Liu Q, Wilson CL, Sapkota Y, Ehrhardt MJ, Gibson TM, Morton LM, Chanock SJ, Neglia JP, Arnold MA, Michael JR, Gout AM, Mulder HL, Easton J, Bhatia S, Armstrong GT, Zhang J, Delaney A, Hudson MM, Robison LL, Yasui Y, Wang Z. Polygenic Risk Score Improves Risk Stratification and Prediction of Subsequent Thyroid Cancer after Childhood Cancer. Cancer Epidemiol Biomarkers Prev. 2021. Epub 2021/09/02. doi: 10.1158/1055-9965.Epi-21-0448. PubMed PMID: 34465587.
12. Wang Z, Liu Q, Wilson CL, Easton J, Mulder H, Chang TC, Rusch MC, Edmonson MN, Rice SV, Ehrhardt MJ, Howell RM, Kesserwan CA, Wu G, Nichols KE, Downing JR, Hudson MM, Zhang J, Yasui Y, Robison LL. Polygenic Determinants for Subsequent Breast Cancer Risk in Survivors of Childhood Cancer: The St Jude Lifetime Cohort Study (SJLIFE). Clin Cancer Res. 2018;24(24):6230-5. Epub 2018/10/28. doi: 10.1158/1078-0432.Ccr-18-1775. PubMed PMID: 30366939; PMCID: PMC6295266.
13. Allan JM, Wild CP, Rollinson S, Willett EV, Moorman AV, Dovey GJ, Roddam PL, Roman E, Cartwright RA, Morgan GJ. Polymorphism in glutathione S-transferase P1 is associated with susceptibility to

chemotherapy-induced leukemia. Proc Natl Acad Sci U S A. 2001;98(20):11592-7. Epub 2001/09/13. doi: 10.1073/pnas.191211198. PubMed PMID: 11553769; PMCID: PMC58774.
14. Imgruet MK, Lutze J, An N, Hu B, Khan S, Kurkewich J, Martinez TC, Wolfgeher D, Gurbuxani SK, Kron SJ, McNerney ME. Loss of a 7q gene, CUX1, disrupts epigenetically driven DNA repair and drives therapy-related myeloid neoplasms. Blood. 2021;138(9):790-805. Epub 2021/09/03. doi: 10.1182/blood.2020009195. PubMed PMID: 34473231; PMCID: PMC8414261.
15. Qin N, Wang Z, Liu Q, Song N, Wilson CL, Ehrhardt MJ, Shelton K, Easton J, Mulder H, Kennetz D, Edmonson MN, Rusch MC, Downing JR, Hudson MM, Nichols KE, Zhang J, Robison LL, Yasui Y. Pathogenic Germline Mutations in DNA Repair Genes in Combination With Cancer Treatment Exposures and Risk of Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. J Clin Oncol. 2020;38(24):2728-40. doi: 10.1200/JCO.19.02760.
16. Richard MA, Lupo PJ, Morton LM, Yasui YA, Sapkota YA, Arnold MA, Aubert G, Neglia JP, Turcotte LM, Leisenring WM, Sampson JN, Chanock SJ, Hudson MM, Armstrong GT, Robison LL, Bhatia S, Gramatges MM. Genetic variation in POT1 and risk of thyroid subsequent malignant neoplasm: A report from the Childhood Cancer Survivor Study. PLoS One. 2020;15(2):e0228887. Epub 2020/02/11. doi: 10.1371/journal.pone.0228887. PubMed PMID: 32040538; PMCID: PMC7010302 following competing interests: Geraldine Aubert is a paid employee of Repeat Diagnostics, Inc., which conducted the telomere length assessment. we can affirm that the employment of Dr. Aubert does not alter our adherence to PLOS ONE policies on sharing data and materials.
17. Green DM, Nolan VG, Goodman PJ, Whitton JA, Srivastava D, Leisenring WM, Neglia JP, Sklar CA, Kaste SC, Hudson MM, Diller LR, Stovall M, Donaldson SS, Robison LL. The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: a report from the Childhood Cancer Survivor Study. Pediatr Blood Cancer. 2014;61(1):53-67. Epub 20130812. doi: 10.1002/pbc.24679. PubMed PMID: 23940101; PMCID: PMC3933293.
18. Morton LM, Sampson JN, Armstrong GT, Chen T-H, Hudson MM, Karlins E, Dagnall CL, Li SA, Wilson CL, Srivastava DK, Liu W, Kang G, Oeffinger KC, Henderson TO, Moskowitz CS, Gibson TM, Merino DM, Wong JR, Hammond S, Neglia JP, Turcotte LM, Miller J, Bowen L, Wheeler WA, Leisenring WM, Whitton JA, Burdette L, Chung C, Hicks BD, Jones K, Machiela MJ, Vogt A, Wang Z, Yeager M, Neale G, Lear M, Strong LC, Yasui Y, Stovall M, Weathers RE, Smith SA, Howell R, Davies SM, Radloff GA, Onel K, de González A B, Inskip PD, Rajaraman P, Fraumeni JJF, Bhatia S, Chanock SJ, Tucker MA, Robison LL. Genome-Wide Association Study to Identify Susceptibility Loci That Modify Radiation-Related Risk for Breast Cancer After Childhood Cancer. J Natl Cancer Inst. 2017;109(11):djx058-djx. doi: 10.1093/jnci/djx058.
19. Seshan VE, Gonen M, Begg CB. Comparing ROC curves derived from regression models. Stat Med. 2013;32(9):1483-93. Epub 2012/10/05. doi: 10.1002/sim.5648. PubMed PMID: 23034816; PMCID: PMC3617074.