Title: Transcriptome-wide association studies reveal key genetic contributors to subsequent neoplasm risk in childhood cancer survivors: a report from the St. Jude Lifetime Cohort and the Childhood Cancer Survivor Study

Authors: Michael J. Betti[1], Chang Li[1], Kendrick Li, Jennifer N French[1], Kateryna Petrykey[1], Kevin Krull[2], Stephanie B. Dixon[1,9], Bonnie Ky[3], Matthew J. Ehrhardt[1,9], Heather M. Conklin[2], Lisa Jacola[2], Angela Delaney[1,4], M. Monica Gramatges[5], Eric J. Chow[6], Kevin Oeffinger[7], ~~Joe~~ Joseph Neglia[8], Lucie M. Turcotte[8], Kirsten K. Ness[1], Melissa M. Hudson[1,9], Gregory T. Armstrong[1], Yutaka Yasui[1], Yadav Sapkota[1]

[1]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA

[2]Department of Psychology and Behavioral Sciences, St. Jude Children's Research Hospital, Memphis, TN, USA

[3]Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4]Department of Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN, USA

[5]Department of Pediatrics, Division of Hematology-Oncology, Baylor College of Medicine, Houston, TX, USA

[6]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA

[7]Department of Medicine, Duke University School of Medicine, Durham, NC, USA

[8]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, USA

[9]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA

Abstract

Background: Subsequent neoplasms (SNs) are among the most common treatment-related adverse effects among survivors of childhood cancer, with an estimated 30-year incidence of 20.5% for any SN and 7.9% for a subsequent malignant neoplasm (SMN). Prior research has identified specific treatment exposures that increase risk for SNs, and genome-wide association studies have identified genetic variants associated with risk for individual SNs in survivors. However, no study to date has utilized transcriptome wide association studies (TWAS) to evaluate the role of genetically predicted enhancer RNA (eRNA) and gene expression in SN risk among survivors.

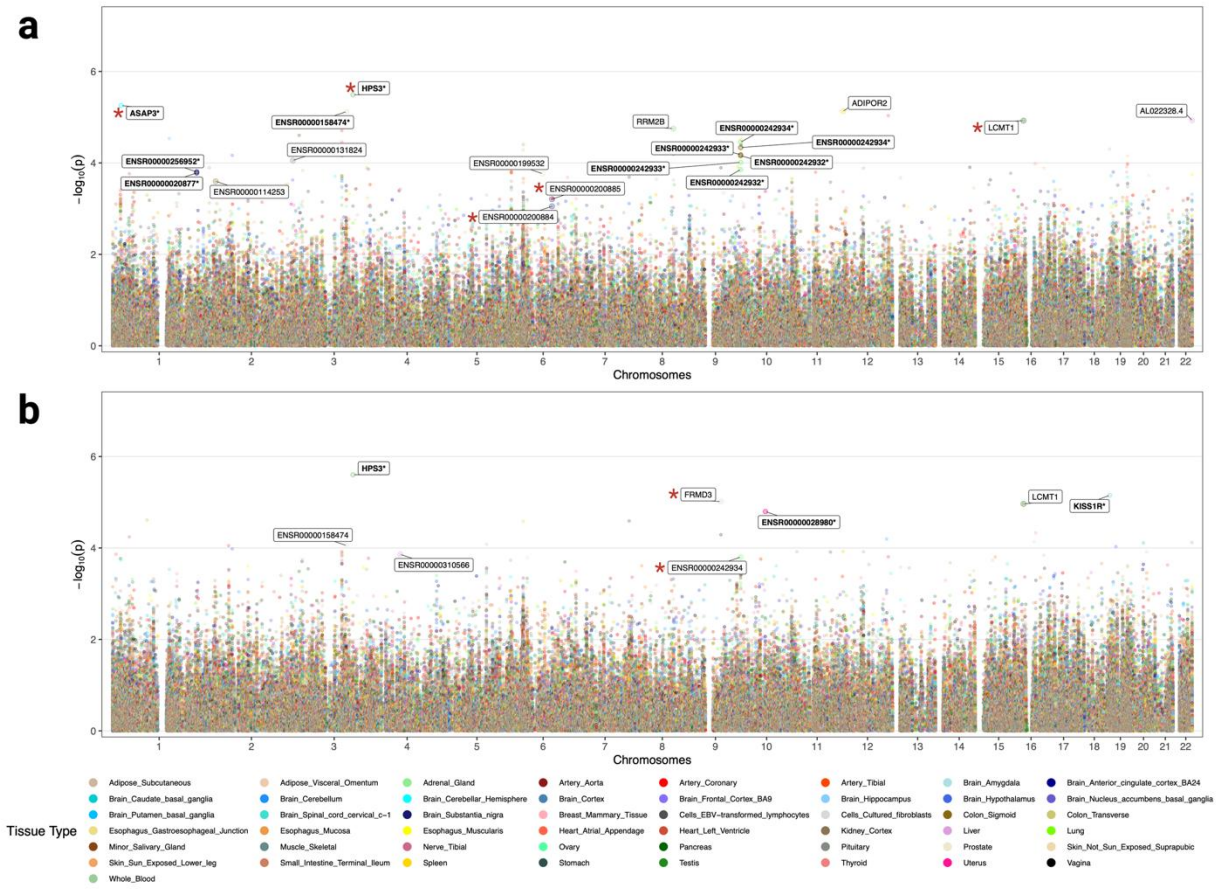
Methods: We applied expression quantitative trait locus (eQTL) models trained in the general population to whole genome sequencing data from the St. Jude Lifetime Cohort Study (SJLIFE) to impute genetically regulated eRNA and gene expression across 49 cell and tissue types. Using predicted expression, we performed logistic regression-based TWAS of SNs (656 cases and 2,569 controls of European ancestry), adjusted for age at primary cancer diagnosis, age at last follow-up, sex, treatment exposures within five years of primary diagnosis (maximum radiation dose to four brain segments, the abdomen, pelvis, chest, and neck, and cumulative doses of alkylating agents, anthracyclines, and epipodophyllotoxins), as well as genetic ancestry. We next performed a secondary TWAS for the subset of SN cases that were

malignant (SMNs: 501 cases and 2,569 controls). Associations with a false discovery rate (FDR)<0.05 were considered significant, while FDR<0.1 was considered suggestive. Significant associations were tested for replication in the Childhood Cancer Survivor Study (CCSS) cohort (SNs: 1,264 cases and 4,725 controls, SMNs: 480 cases and 4,725 controls). Replication was defined as p<0.1 with concordant direction of effect in at least one tissue.

Results: TWAS for SNs revealed 11 eRNA (two replicated) and six gene transcripts (three replicated) showing at least suggestive association. *HPS3* (in whole blood), a gene involved in organelle biogenesis, was the strongest TWAS association for SNs (discovery: OR=1.23, p=3.24x10^{-6}; replication: OR=1.08, p=0.01). Other notable associations with SN risk were oncogene *ASAP3* in the cerebellar hemisphere (discovery: OR=0.86, p=5.6 x 10^{-6}, replication: OR=0.95, p=0.07) and tumor suppressor gene *LCMT1* in pancreas (discovery: OR=1.09, p=1.19x10^{-5}; replication: OR=1.07, p=0.06). In a secondary analysis restricted to individuals with an SMN, we identified four eRNA (one replicated) and four gene (one replicated) associations. In addition to *HSP3* in whole blood (discovery: OR=1.22, p=2.50 x 10^{-6}, replication: OR=1.02, p=0.13) and *LCMT1* in pancreas (discovery: OR=1.09, p=1.08x10^{-5}; replication: OR=1.01, p=0.38), genes that previously associated with risk for SNs, we also discovered an association with tumor suppressor gene *FRMD3* in fibroblasts (discovery: OR=1.23, p=9.6x10^{-6}, replication: OR=1.04, p=0.02).

Conclusions: TWAS identified 17 unique transcripts associated with SNs and eight with SMNs, demonstrating the promise of TWAS for discovering biologically interpretable contributors to secondary cancer risk in childhood cancer survivors. Future work will utilize functional validation experiments to interrogate the causal mechanisms of these TWAS associations. Ultimately, these findings may help to improve risk prediction and identify therapeutic targets for future clinical intervention.

Figure:



TWAS results for subsequent neoplasms (SNs) and subsequent malignant neoplasms (SMNs) in childhood cancer survivors. a, TWAS of SNs across 49 cell and tissue types (656 cases and 2,569 controls from the St. Jude Lifetime Cohort Study). Genome-wide significant associations (FDR<0.05) are labeled in bold, and suggestive associations (FDR<0.1) are labeled in non-bold font. eRNA names begin with the prefix “ENSR.” Associations that replicated in the CCSS are marked with a red asterisk. **b,** TWAS of SMNs (501 cases and 2,569 controls from the St. Jude Lifetime Cohort Study).