

Title: Trans-ancestral genetic study of diabetes mellitus risk in survivors of childhood cancer: a report from the St. Jude Lifetime Cohort and the Childhood Cancer Survivor Study

Authors: Cindy Im,^{1*} Achal Neupane,² Jessica L. Baedke,² Angela Delaney,^{2,3} Stephanie B. Dixon,^{2,4} Eric J. Chow,⁵ Sogol Mostoufi-Moab,⁶ Melissa A. Richard,⁷ M. Monica Gramatges,⁷ Philip J. Lupo,⁷ Noha Sharafeldin,⁸ Smita Bhatia,⁸ Gregory T. Armstrong,^{2,4} Melissa M. Hudson,^{2,4} Kirsten K. Ness,² Leslie L. Robison,² Yutaka Yasui², Carmen L. Wilson,^{2*†} Yadav Sapkota^{2*†}

Affiliations:

1. Department of Pediatrics, University of Minnesota, Minneapolis, MN, 55455, USA
2. Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA
3. Division of Endocrinology, Department of Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA
4. Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA
5. Public Health Sciences and Clinical Research Divisions, Fred Hutchinson Research Center, Seattle, WA, 98109, USA
6. Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, 19146, USA
7. Section of Hematology-Oncology, Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030, USA
8. Institute for Cancer Outcomes and Survivorship, University of Alabama at Birmingham, Birmingham, AL, 35223, USA

†These senior authors contributed equally to this work.

***Corresponding authors:**

Cindy Im (imcindy@umn.edu)
Department of Pediatrics
University of Minnesota
420 Delaware Street SE, MMC 715
Minneapolis, MN, 55455, USA

Carmen Wilson (carmen.wilson@stjude.org)
Department of Epidemiology and Cancer Control
St. Jude Children's Research Hospital
262 Danny Thomas Place, Mail Stop 735
Memphis, TN 38105, USA

Yadav Sapkota (yadav.sapkota@stjude.org)
Department of Epidemiology and Cancer Control
St. Jude Children's Research Hospital
262 Danny Thomas Place, Mail Stop 735
Memphis, TN 38105, USA

Acknowledgements: This study was supported by the US National Cancer Institute (R21 CA261833, C Im/Y Yuan principal investigators; R01 CA216354, Y Yasui/J Zhang principal investigators; U01 CA195547, MM Hudson/K Ness, principal investigators; R01 CA261898, P Burrige/Y Sapkota, principal investigators; U24 CA55727, GT Armstrong, principal investigator; Cancer Center Support [CORE] Grant CA21765, C Roberts, principal investigator), and the American Lebanese Syrian Associated Charities.

ABSTRACT

Type 2 diabetes mellitus (T2D) is an established late effect of treatment for childhood cancer. Leveraging detailed cancer treatment and whole-genome sequencing data among survivors of childhood cancer of European (EUR) and African (AFR) genetic ancestry in the St. Jude Lifetime Cohort (N=3,676; 304 cases), five novel diabetes mellitus (DM) risk loci were identified with independent trans-/within-ancestry replication, including in 5,965 survivors of the Childhood Cancer Survivor Study. Among these, common risk variants at 5p15.2 (*LINC02112*), 2p25.3 (*MYT1L*), and 19p12 (*ZNF492*) modified alkylating agent-related risks across ancestry groups, but AFR survivors with risk alleles experienced disproportionately greater risk of DM (AFR, variant ORs: 3.95-17.81; EUR, variant ORs: 2.37-3.32). Novel risk locus *XNDC1N* was identified in the first genome-wide DM rare variant burden association analysis in survivors (OR=8.65, 95% CI: 3.02-24.74, $P=8.1 \times 10^{-6}$). Lastly, a general-population 338-variant multi-ancestry T2D polygenic risk score was informative for DM risk in AFR survivors, and showed elevated DM odds after alkylating agent exposures (quintiles: combined $OR_{EUR}=8.43$, $P=1.1 \times 10^{-8}$; $OR_{AFR}=13.85$, $P=0.033$). This study supports future precision diabetes surveillance/survivorship care for all childhood cancer survivors, including those with AFR ancestry.

MAIN TEXT

With advances in therapy, five-year survival after childhood cancer diagnosis has improved to over 85%¹. However, as a consequence of exposures to these curative therapies, long-term (≥ 5 years) survivors face excess chronic disease morbidity²⁻⁴, including increased risk for developing type 2 diabetes mellitus (T2D)^{5,6} later in life. Survivors demonstrate an almost 3-fold greater risk for diabetes mellitus (DM), primarily T2D^{5,6}, compared to siblings^{5,7}. The risk for DM is elevated among survivors who received abdominal radiation therapy (RT) or total body irradiation, and to a lesser extent, alkylating agents^{5,6,8,9}. Additionally, younger ages at primary cancer diagnosis, increasing age, and higher body mass index (BMI) also contribute to risk^{6,9}.

Known clinical risk factors do not fully explain why some survivors develop DM while others do not. Genetic susceptibility is posited to play a role. Genome-wide association studies (GWAS), including large-scale meta-analyses with >1 million participants in the general population, have identified >550 loci associated with T2D risk, including those shared across diverse ancestral groups^{10,11}. Analyses of rare variants (minor allele frequencies $<1\%$) have revealed additional T2D risk loci¹² and have been shown to enhance T2D risk stratification¹³. But given that DM genetic susceptibility among survivors is anticipated to additionally reflect the risk contributions conferred by cancer treatments, novel investigations of DM genetic risk are also needed among survivors of childhood cancer¹⁴. A recent GWAS of DM risk among 5,083 survivors of European ancestry participating in the Childhood Cancer Survivor Study (CCSS) in which cases were identified through self-report identified one novel genome-wide significant risk locus at 9q22.32 (*ERCC6L2*)¹⁵.

To date, our understanding of DM genetic susceptibility factors among childhood cancer survivors are limited while studies including survivors of diverse ancestry or exploring the contribution of rare variants to DM risk do not exist, to our knowledge. Hence, we report findings from a trans-ancestry GWAS and a gene-based rare variant burden analysis with whole-

genome sequencing (WGS) data and clinically-ascertained DM status in 3,676 survivors of European and African genetic ancestry participating in the St. Jude Lifetime Cohort (SJLIFE). With the identification of five novel DM risk loci, we highlight the need to study common and rare genetic susceptibility factors contributing to risk of late effects among diverse populations of childhood cancer survivors to provide future precision surveillance/survivorship care.

RESULTS

The demographic and clinical characteristics of SJLIFE childhood cancer survivors included in trans-ancestral and rare variant discovery analyses are provided in Table 1. A total of 261 (8.4%) and 43 (7.5%) survivors of European genetic ancestry (EUR; N=3,102) and African genetic ancestry (AFR; N=574) developed DM, respectively. SJLIFE AFR survivors were comparatively younger at follow up (median=28 years, IQR: 21-37 years) than SJLIFE EUR survivors (median=32 years, IQR: 24-41 years). The prevalence of overweight/obesity (BMI >25 kg/m²) was similar between ancestry groups (~60%). Nearly one-fifth of SJLIFE survivors received abdominal RT, while 57% and 52% of EUR and AFR survivors were treated with alkylating agents, respectively. Non-genetic covariates used in downstream genetic analyses were the same as those used in a previous DM GWAS in CCSS¹⁵, with abdominal RT and alkylating agents as cancer treatment risk factors of interest (Supplemental Table 1). BMI was significantly associated with DM risk in SJLIFE survivors (EUR: OR=1.56 per 5 kg/m², P=3.8x10⁻²⁸; AFR: OR=1.45 per 5 kg/m², P=3.9x10⁻⁴), followed by abdominal RT dose (OR=1.23 per 10 Gray, P=3.6x10⁻⁴) and any treatment with alkylating agents (OR=1.35, 95% CI: 1.02-1.80, P=0.039) among EUR survivors. For the replication study conducted in EUR survivors participating in the Childhood Cancer Survivor Study (CCSS), 223 (5.7%) and 47 (2.3%) self-reported use of glucose-lowering medications to treat DM in the Original (cancer diagnosed

1970-1986) and Expansion (cancer diagnosed 1986-1999) cohorts, respectively (Supplementary Table 2).

Common genetic variants with trans-ancestry associations with DM risk

We conducted ancestry-specific GWAS for clinically-ascertained DM in SJLIFE with logistic regression models adjusted for relevant clinical and sociodemographic risk factors, separately analyzing 3,102 EUR and 574 AFR survivors. To facilitate discovery of common loci with trans-ancestral DM risk effects in survivors, we prioritized all variants within a 100-kb window of the most significantly associated index variant ($P < 5 \times 10^{-6}$) with effect allele frequencies (EAFs) $\geq 1\%$ in both SJLIFE EUR and AFR samples for further investigation, using trans-ancestral replication (i.e., replication between ancestry groups, with a Bonferroni-corrected p-value threshold) and trans-ancestry meta-analysis to declare genome-wide significance ($P < 5 \times 10^{-8}$) (see Methods). We then utilized available data from CCSS EUR survivors for replication of GWAS findings that were genome-wide significant in the EUR-specific GWAS, and for secondary replication of genome-wide significant trans-ancestry meta-analysis findings. Replication in a second sample of AFR survivors was not feasible in CCSS due to sample size limitations (N=149, with 6 cases).

Genetic variants with replicated trans-ancestral and ancestry-specific DM risk effects are shown in Table 2 (Manhattan and quantile-quantile plots from SJLIFE ancestry-specific GWAS provided in Supplementary Figures 1 and 2). In the EUR-specific GWAS in SJLIFE, we identified one genome-wide significant locus on 8q11.21 within *SNTG1* (syntrophin-gamma 1) (Supplementary Figure 1). The most significantly associated variant at 8q11.21 in the SJLIFE EUR discovery dataset was rs6993752 (OR=1.99, 95% CI: 1.56-2.55; $P=4.4 \times 10^{-8}$) (Table 2, Supplementary Table 3), replicating in CCSS EUR survivors (OR=1.56, 95% CI: 1.12-2.17; $P=8.1 \times 10^{-3}$) (Table 2, Supplementary Tables 4, 5). In SJLIFE AFR survivors, these 8q11.21 variants were associated with DM ($P < 0.05$) but showed ancestral allelic effect heterogeneity

(Supplementary Table 3), which can occur due to differences in LD with causal variants, environment, or polygenic background across ancestrally diverse populations¹⁶. Effect allele frequencies for 8q11.21 variants were generally higher in SJLIFE AFR survivors, but showed a broken pattern of LD in the AFR 1000 Genomes reference panel relative to rs6993752 which was not observed in EUR samples (Supplementary Figure 4).

We also prioritized two loci with suggestive associations with DM risk ($P < 5 \times 10^{-6}$) identified among SJLIFE EUR survivors: 2p25.3, represented by three variants upstream of *MYT1L* (myelin transcription factor 1 like; most significantly associated [index] variant rs143578296: OR=1.99, 95% CI: 1.49-2.66; $P = 2.8 \times 10^{-6}$) and 19p12, represented by five variants upstream of *ZNF492* (zinc finger protein 492; index variant rs35017858: OR=2.26, 95% CI: 1.60-3.20; $P = 4.1 \times 10^{-6}$) (Table 2, Figure 1). We found that all variants at the 2p25.3 and 19p12 loci were replicated in SJLIFE AFR survivors (Supplementary Table 3) at nominal significance ($P < 0.05$), while the locus index SNPs were replicated under a Bonferroni-corrected p-value threshold that accounted for the number of tested SNPs. The index SNPs at 2p25.3 and 19p12 attained genome-wide significance in trans-ancestry meta-analysis (2p25.3, rs143578296: OR=2.05, 95% CI: 1.59-2.65, $P = 4.5 \times 10^{-8}$; 19p12, rs35017858: OR=2.43, 95% CI: 1.80-3.28; $P = 5.7 \times 10^{-9}$). Variants at the 2p25.3 locus were also replicated in CCSS EUR survivors (rs55726572 as the best LD proxy, OR=1.45, 95% CI: 1.00-2.10, $P = 0.047$).

In the AFR-specific GWAS in SJLIFE, we identified a novel variant at the 5p15.2 locus (Supplementary Figure 5) with suggestive significance (rs13436710: OR=10.19, 95% CI: 4.12-25.21, $P = 5.1 \times 10^{-7}$). While variants in LD with rs13436710 were low-frequency in EUR samples (EAF ~1%), index/proxy variants replicated in both SJLIFE EUR (OR=2.47, 95% CI: 1.23-4.98, $P = 0.011$) and CCSS EUR (OR=1.48, 95% CI: 1.06-2.07, $P = 0.021$) survivors.

We examined associations between all available replicated common index variants and self-reported and ICD-coded T2D that were tested in the UK Biobank Study (UKBB; $N > 361,000$) (Supplementary Table 6). The T2D risk effect sizes in the UKBB were extremely modest

(OR<1.002) for the index variant at the 2p25.3 locus (rs143578296), which showed nominal association with ICD-10-coded T2D risk ($P=3.6 \times 10^{-3}$), and the index variant at the 19p12 locus (rs35017858), which showed nominal association with self-reported T2D risk ($P=6.6 \times 10^{-3}$).

Allelic risks for developing DM and the contributions of cancer treatments and adiposity

We investigated whether replicated DM risk variants from common variant analyses in survivors potentially modified DM risk effects following exposures to specific cancer treatments. While risk alleles at the 2p25.3, 5p15.2, and 19p12 loci did not appear to modify abdominal RT-related DM risk effects in both SJLIFE EUR and AFR survivors (Supplementary Table 7), index variants at these loci were associated with appreciably greater DM risks among SJLIFE EUR and AFR survivors treated with alkylating agents, i.e., alkylating agent-related risk effect modifications seen in the discovery data were replicated in another ancestral dataset (Figure 1, Supplementary Table 8). The magnitude of DM associations in SJLIFE AFR survivors treated with alkylating agents was greater for all variants compared to those seen in SJLIFE EUR survivors treated with alkylating agents. For example, the 19p12 index DM risk allele (rs35017858) identified in SJLIFE EUR survivors was associated with 3.3-fold greater DM odds in SJLIFE EUR survivors with high-dose alkylating agent exposures (95% CI: 1.91-5.79, $P=2.3 \times 10^{-5}$), but 5.7-fold greater DM odds in the replication dataset of SJLIFE AFR survivors with high-dose alkylating agent exposures (95% CI: 1.99-16.54, $P=1.2 \times 10^{-3}$). Corresponding allelic risks among survivors without this treatment exposure was appreciably lower (EUR: OR=1.63, $P=0.22$; AFR: OR=2.00, $P=0.096$). From the SJLIFE AFR discovery GWAS, the 5p15.2 index variant (rs13436710) risk allele was associated with 17.8-fold (95% CI: 3.46-91.60, $P=5.7 \times 10^{-4}$) greater odds of DM in alkylating agent-exposed AFR survivors, in contrast to the 3.3-fold (95% CI: 1.44-7.68, $P=5.0 \times 10^{-3}$) greater DM odds observed in the replication dataset of SJLIFE EUR survivors. In comparison, corresponding allelic risks among survivors without this treatment exposure were lower (EUR: OR=1.27, $P=0.73$; AFR: OR=10.84, $P=2.1 \times 10^{-4}$).

The index variant at the 8q11.21 locus (rs6993752) did not show evidence of modifying DM risks related to abdominal RT or alkylating agents. Given treatment with abdominal RT has been previously linked to lower adiposity among survivors¹⁷ but obesity is strongly associated with other cancer treatments and DM risk⁶, we evaluated whether this DM risk allele showed significant effect heterogeneity by adiposity. Overweight/obese (BMI>25 kg/m²) EUR survivors in SJLIFE were significantly more likely to have received cranial RT (P<0.001) or corticosteroids (P<0.001). We found that DM risks associated with the risk allele at the 8q11.21 locus index variant (rs6993752) were appreciably greater among overweight/obese EUR survivors in SJLIFE (OR_{BMI>25}=2.23, 95% CI: 1.72-2.89, P=1.8x10⁻⁹ versus OR_{BMI≤25}=0.90, P=0.80) and CCSS (OR_{BMI>25}=1.69, 95% CI: 1.15-2.47, P=7.0x10⁻³ versus OR_{BMI≤25}=1.18, P=0.63) (Figure 1, Supplementary Table 9).

T2D polygenic risk in survivors

We assessed whether T2D PRSs derived in non-cancer study populations were associated with DM risk in childhood cancer survivors, and in particular, whether a 338-variant T2D PRS based on general population multi-ancestry GWAS meta-analyses¹¹ also had increased transferability in survivors of AFR ancestry. We found standard deviation increases in the multi-ancestry T2D PRS were significantly associated with increased odds of DM in all SJLIFE and CCSS survivor cohorts (SJLIFE AFR: OR=1.80, P=2.8x10⁻³; SJLIFE/CCSS EUR: ORs=1.57-1.84, P<9.5x10⁻⁴) (Table 3). While a ~6.9 million-variant PRS derived from a T2D GWAS in Europeans¹⁸ was also associated with DM in all SJLIFE and CCSS EUR survivor cohorts (ORs: 2.16-3.06, P<6.8x10⁻⁴), this EUR-only PRS was not associated with DM in SJLIFE AFR survivors (OR=0.97, P=0.95).

The 338-variant multi-ancestry T2D PRS¹¹ was further evaluated for potential treatment-related DM risk effect modification (Figure 2, Supplementary Table 10). No consistent evidence of variation in the odds of DM by abdominal RT exposures was observed (top vs. lowest PRS

quintile in SJLIFE and CCSS EUR survivors with no abdominal RT exposures: $OR_{Q5}=4.90$, $P=3.6 \times 10^{-13}$, versus with abdominal RT exposures ≥ 20 Gy: $OR_{Q5}=3.56$, $P=2.5 \times 10^{-5}$). However, the multi-ancestry T2D PRS was consistently associated with increasing odds of DM among EUR and AFR survivors treated with alkylating agents (Figure 2). The top T2D PRS quintile was associated with 4.1-fold greater odds of DM (95% CI: 2.16-7.69; $P=1.5 \times 10^{-5}$) in SJLIFE and CCSS EUR survivors who were not treated with alkylating agents, whereas an 8.4-fold greater odds of DM (95% CI: 4.06-17.53; $P=1.1 \times 10^{-8}$) was observed in those with the same level of polygenic risk but were treated with alkylating agent doses $\geq 4,000$ mg/m². A similar trend was seen in SJLIFE AFR survivors, where those in the top T2D PRS quintile who were treated with any alkylating agents had 13.9-fold greater DM odds (95% CI: 1.24-154.48, $P=0.033$) than those in the lowest.

Rare pathogenic variant burden in the *XNDC1N* gene and DM risk

For genome-wide gene-based rare variant burden analyses using pathogenic/likely pathogenic (P/LP) variants, we randomly split the SJLIFE EUR cohort *a priori* into discovery (N=1,550) and replication (N=1,552) datasets (see Methods). In the rare variant association analysis, we identified two genes with P/LP variant burden enrichment suggestively associated with DM in the SJLIFE EUR discovery sample (N=1,550; $P < 1.0 \times 10^{-3}$): *XNDC1N* (XRCC1 N-terminal domain containing 1, N-terminal like; burden test $P=8.5 \times 10^{-4}$) and *ZNF655* (zinc finger protein 655; burden test $P=7.1 \times 10^{-4}$) (Supplementary Table 11). In the separate replication analysis (N=1,552), only *XNDC1N* remained significantly associated ($P=7.1 \times 10^{-4}$). In the combined data, the association between DM risk and *XNDC1N* P/LP burden was genome-wide significant (Bonferroni-corrected $P < 0.05/679$ genes = 7.4×10^{-5}). Carrying either of the two *XNDC1N* P/LP variants conferred 8.65 greater odds of DM risk (MAF=0.003, 95% CI: 3.02-24.74; burden test $P=8.1 \times 10^{-6}$) (Table 4). Single variant analyses further revealed that both *XNDC1N* P/LP variants contribute to increased DM risk (Supplementary Table 12). For

secondary validation, we compared *XNDC1N* rare pathogenic risk allele frequencies among healthy EUR adults in the Genome Aggregation Database (gnomAD) to SJLIFE EUR survivor controls and cases (Supplementary Table 13). Compared to healthy gnomAD controls, the odds of carrying *XNDC1N* risk alleles among SJLIFE EUR controls were similar. In contrast, the odds of carrying these risk alleles were significantly higher among survivors with DM compared to healthy gnomAD controls (rs200142185: OR=4.72, 95% CI: 1.33-13.94, $P=8.8 \times 10^{-3}$; rs771623613, OR=19.77, 95% CI: 2.26-236.43, $P=3.2 \times 10^{-3}$).

***In silico* functional annotation of DM risk variants**

Functional annotations for all novel DM risk variants are provided in Supplementary Tables 14 and 15. The DM risk allele at the 8q11.21 locus index variant (rs6993752) is associated with *SNTG1*-region hypermethylation in BIOS QTL¹⁹ and overlaps putative promoter and enhancer chromatin states in liver and pancreas epigenomes. Index variants at 2p25.3, 5p15.2, and 19p12, all with evidence of trans-ancestral DM risk effects, overlap regions annotated with a Polycomb-repressed chromatin state in pancreatic cells. The 5p15.2 index variant risk allele is associated with increased alternative *FAM173B* splicing²⁰, while the 19p12 index variant risk allele is associated with decreased transcription of nearby zinc finger proteins (*ZNF257*, *ZNF492*) in multiple tissues²⁰. The *XNDC1N* rare pathogenic alleles are protein-truncating variants, overlapping transcribed regions in fat, liver, and pancreas epigenomes.

DISCUSSION

Using a well-characterized cohort of long-term survivors of childhood cancer with clinically-ascertained DM status and WGS data, we identified five novel DM risk loci. Among these, we found that common risk variants at three loci, 5p15.2 (*LINC02112*), 2p25.3 (*MYT1L*), 19p12 (*ZNF492*), contributed to increased risk of DM in both EUR and AFR survivors treated

with alkylating agents. In particular, AFR survivors with these DM risk alleles who were treated with alkylating agents experienced disproportionately greater odds of DM compared to their EUR counterpart. We also confirmed a general population 338-variant T2D PRS developed from a multi-ancestry genetic meta-analysis¹¹ was potentially useful to survivors of AFR ancestry. Interestingly, this multi-ancestry T2D PRS was also associated with strikingly consistent increases in DM odds among EUR and AFR survivors treated with alkylating agents, contributing to its potential clinical utility for DM risk assessment in both ancestry groups. This study provides a step forward towards informing precision diabetes surveillance and survivorship care for all childhood cancer survivors, including those in minority ancestry groups.

Genetic and epigenetic regulation plausibly contribute to both inter- and intra-individual variation in cellular responses to alkylating agents^{21,22}, suggesting DM susceptibility can be influenced by both variations in genetic risk and exposures to alkylating agents. Notably, all three trans-ancestral DM risk loci identified in this study overlap putative Polycomb-repressed chromatin states in pancreatic cells. Polycomb group proteins have been shown to be involved in the epigenetic regulation of pancreatic β -cell replication, where the loss of Polycomb complex binding in pancreatic β -cells has been linked to reduced replication activity²³. While it is unclear how the risk alleles identified in this study explicitly contribute to DM pathophysiology, these loci are strong candidates for further investigation. For example, rare deletions on the short arm of chromosome 2 (2p25.3) spanning the *MYT1L* gene have previously been linked to early-onset obesity and intellectual disability^{24,25} and DM²⁶. The 5p15.2 index variant is mapped to an intronic region of *LINC02112*. *LINC02112* includes variants suggestively associated with cytokine concentrations (interleukin-2 receptor antagonists), which have been linked to inflammatory diseases including DM²⁷. Based on GTEx²⁰ data, the 19p12 locus index variant is significantly associated with decreased expression of *ZNF257*. *ZNF257* variants common in East Asian populations but which are rare/monomorphic in European populations have been linked to T2D risk in a previous GWAS of 433,540 individuals of East Asian genetic ancestry²⁸.

Interestingly, we also observed increased risk for DM associated with common variants at the 8q11.21 locus among overweight/obese EUR childhood cancer survivors. The index variant (rs6993752) did not appear to modify abdominal RT- or alkylating agent-related DM risks. This finding suggests DM-associated variants at 8q11.21 could increase risk through adiposity-related pathways, involving any of a myriad of risk factors for obesity, including exposures to cranial RT and glucocorticoids. The 8q11.21 index variant is located in an intronic region in *SNTG1* and is a methylation quantitative trait loci for this gene in whole blood in BIOS QTL¹⁹. While *SNTG1* has not been associated with DM risk in previous GWAS, other studies have identified *SNTG1* as a blood pressure locus²⁹, a phenotype strongly linked to obesity.

We conducted the first genome-wide rare variant burden association analysis for DM in survivors, leading to the identification of *XNDC1N* (XRCC1 N-terminal domain containing 1) as a novel potential DM risk locus. These results were further supported by comparisons of *XNDC1N* risk allele frequencies between survivor controls and cases and healthy general population (gnomAD) adults. *XNDC1N* is predicted to encode XRCC1 N-terminal domain-like proteins, which have been shown contribute to XRCC1 protein function and the DNA repair process³⁰. Specifically, XRCC1 N-terminal domain proteins modulate XRCC1 activity and are directly involved in the repair of single-strand DNA breaks, preferentially binding to single-strand break DNA^{31,32}. There is evidence that mutations in the *XRCC1* gene in cell lines show characteristics of DNA repair defects consistent with exposures to ionizing radiation, methylating reagents, and alkylating agents^{33,34}. While mutations in *XNDC1N* have not been linked to radiation or chemotherapy exposures in human populations, polymorphisms in *XRCC1* have been associated with radiation-induced late toxicity among breast cancer³⁵ and prostate cancer patients³⁶. Given the *XNDC1N* locus is putatively transcribed in fat, liver, and pancreas epigenomes, protein-truncating variants in *XNDC1N* may plausibly contribute to DM susceptibility among survivors.

The main strengths of our analysis include clinical ascertainment of DM status, rather than self-report, among SJLIFE participants, inclusion of a non-European genetic ancestry survivor population in analyses, and genome-wide interrogation of rare P/LP burden and DM risk. Despite these strengths, our analysis had several limitations. First, both our common and rare variant analyses are limited by the lack of secondary replication in independent cohorts of survivors with clinically ascertained DM status and WGS-based genotype data with sufficient sample size. However, such datasets do not currently exist. To minimize the possibility of false positive findings, we implemented study designs with separate replication studies, i.e., replication between ancestry subgroups in SJLIFE and secondary replication in CCSS for common variant GWAS, and an *a priori* discovery-replication data split for rare variant analysis. Second, our analyses have limited statistical power to formally test for gene-treatment statistical interactions. Top associations from variant main effects analyses may not overlap with top findings from a genome-wide analysis of gene-treatment interactions. However, the plausibility of treatment- and adiposity-related DM risk effects was supported by replicating stratified results across multiple study samples. Finally, *in silico* functional annotation is not a substitute for experimental functional validation. Thus, biological follow-up of these findings to elucidate their contributions to DM pathophysiology among childhood cancer survivors is warranted.

In summary, we identified five novel risk loci for DM, including a novel gene enriched with rare DM-associated protein-truncating variants, among childhood cancer survivors. The DM susceptibility conferred by these loci appeared to be modulated in part by survivors' cancer treatment exposures. Trans-ancestry genetic analyses revealed three of these novel DM loci were associated with disproportionately greater alkylating agent-related risk among African-ancestry survivors. We also observed that a general population T2D PRS based on analyses in European samples did not appear to be informative for assessing DM risks in African-ancestry survivors, but a general population multi-ancestry T2D PRS may be relevant to more diverse populations of survivors, which may in part be due to its potential modification of alkylating

agent-related DM risks across ancestral groups. Further investigations are needed to understand how these genetic risk factors contribute to DM risk mechanisms and can be utilized to support DM risk surveillance among survivors of diverse ancestry.

TABLES

Table 1: Demographic and clinical characteristics of childhood cancer survivors in the SJLIFE cohort, by ancestry

Characteristic	SJLIFE, European ancestry (N=3102)		SJLIFE African ancestry (N=574)	
	N	%	N	%
Sex				
Male	1648	53.1	286	49.8
Female	1454	46.9	288	50.2
Age at follow up (in years)				
<25	904	29.1	228	39.7
25-34	957	30.9	179	31.2
35-44	731	23.6	107	18.6
≥45	510	16.4	60	10.5
Body mass index ^a (kg/m ²)				
≤25	1159	37.4	217	37.8
>25 to ≤30	822	26.5	133	23.2
>30 to ≤40	899	29.0	163	28.4
>40	222	7.2	61	10.6
Developed type 2 diabetes	261	8.4	43	7.5
Age at cancer diagnosis (in years)				
<5	1323	42.7	244	42.5
5-9	672	21.7	119	20.7
10-14	630	20.3	129	22.5
≥15	477	15.4	82	14.3
Childhood cancer diagnosis				
Leukemias	1040	33.5	115	20.0
Central nervous system tumors	436	14.1	81	14.1
Hodgkin lymphoma	343	11.1	60	10.5
Non-Hodgkin lymphoma	203	6.5	43	7.5
Kidney tumors	184	5.9	62	10.8
Soft tissue sarcomas	193	6.2	52	9.1
Other	703	22.7	161	28.1
Cancer treatments				
Any abdominal radiation therapy (RT)	566	18.3	106	18.5
Any alkylating agents	1751	56.5	299	52.1
Dosages: Median (IQR) ^b				
Abdominal RT tumor dose (cGy)	566	2400 (2000-3500)	106	2050 (1200-3375)
Alkylating agents (CED, mg/m ²)	1751	8155 (4425-13271)	299	7448 (3696-11783)

Abbreviations: SJLIFE, St. Jude Lifetime Cohort; IQR, interquartile range; RT, radiation therapy; CED, cyclophosphamide equivalent dose; mg, milligrams; m, meters; kg, kilograms.

a. Body mass index values were adjusted for amputation.

b. Treatment dose distribution statistics are provided for childhood cancer survivors who received any of the specified treatments and had non-missing dosage information.

Table 2: Ancestry-specific and trans-ancestral diabetes risk variants in childhood cancer survivors

Chr	BP	EA	NEA	rsid (nearest gene)	Ancestry group	SJLIFE ^a (cancer treated 1962-2012; N=3102 EUR; N=574 AFR)			CCSS original cohort ^b (cancer diagnosed 1970-1986; N=3902 EUR)			CCSS expansion cohort ^b (cancer diagnosed 1987-1999; N=2063 EUR)		
						EAF	OR (95% CI)	P	EAF	OR (95% CI)	P	EAF	OR (95% CI)	P
SJLIFE EUR survivors: Top associations ($P < 5 \times 10^{-6}$), within-ancestry replication ($P < 0.05$)														
8	50638825	G	A	rs6993752 (<i>SNTG1</i>)	EUR	0.15	1.99 (1.56-2.55)	4.4×10^{-8}	0.16	1.56 (1.12-2.17)	8.1×10^{-3}	0.15	0.52 (0.22-1.20)	0.13
					AFR	0.38	0.53 (0.30-0.91)	0.022						
					EUR + AFR meta-analysis ^c		+-	1.4×10^{-7}						
SJLIFE EUR survivors: Top associations ($P < 5 \times 10^{-6}$), trans-ancestral replication ($P < 0.05$)														
2	2434309	A	AAAATAAAATAAA ATAAAATAAAAC	rs143578296 ^d (<i>MYT1L</i>)	EUR	0.10	1.99 (1.49-2.66)	2.8×10^{-6}	0.14	1.45 (1.00-2.10)	0.047	0.13	0.91 (0.38-2.14)	0.82
					AFR	0.17	2.30 (1.30-4.07)	4.2×10^{-3}						
					EUR + AFR meta-analysis		2.05 (1.59-2.65)	4.5×10^{-8}						
19	22614854	G	A	rs35017858 (<i>LINC01785</i> ; <i>ZNF257</i> ; <i>ZNF492</i>)	EUR	0.05	2.26 (1.60-3.20)	4.1×10^{-6}	0.06	1.04 (0.57-1.89)	0.91	0.06	0.75 (0.22-2.60)	0.65
					AFR	0.22	2.99 (1.66-5.37)	2.6×10^{-4}						
					EUR + AFR meta-analysis		2.43 (1.80-3.28)	5.7×10^{-9}						
SJLIFE AFR survivors: Top associations ($P < 5 \times 10^{-6}$) with trans-ancestral replication ($P < 0.05$)														
5	9843018	C	T	rs13436710 ^e (<i>LINC02112</i>)	AFR	0.05	10.19 (4.12-25.21)	5.1×10^{-7}	0.07	1.48 (1.06-2.07)	0.021	0.01	1.02 (0.13-7.77)	0.98
					EUR	0.01	2.47 (1.23-4.98)	0.011						
					AFR + EUR meta-analysis ^c		++	2.0×10^{-7}						

Abbreviations: EUR, European genetic ancestry; AFR, African genetic ancestry; Chr, chromosome; BP, base position, GRCh38 (hg38) build; EA, effect (risk) allele; NEA, non-effect (reference) allele; rsid, genetic variant identifier using dbSNP build 151; EAF, effect allele frequency in analytic sample; OR, odds ratio; CI, confidence interval; P, p-value.

a. Associations with $P < 5 \times 10^{-6}$ in the SJLIFE discovery GWAS were prioritized for replication studies.

b. For index variants identified in EUR SJLIFE discovery GWAS, CCSS replication results for self-reported diabetes mellitus requiring insulin are reported. The EAF of the top AFR index variant in the 1000 Genomes EUR population is low, therefore replication in CCSS for self-reported diabetes mellitus requiring oral medications or insulin is reported.

c. Due to evidence of ancestral allelic heterogeneity ($P_{\text{het}} < 0.05$), meta-analysis p-values were estimated using the Han-Eskin random-effects model (RE2). Sample directions of effect are shown (discovery followed by replication).

d. Association test results in CCSS are shown for the best available LD proxy in EUR 1000 Genomes population (rs55726572; chr2:2460867:T>C; $D' = 0.92$, $R^2 = 0.64$).

e. Association test results in the CCSS original cohort are shown for the best available LD proxy in EUR 1000 Genomes (rs10475476; chr5:9841183:C>G; $D' = 1$, $R^2 = 0.0998$).

Table 3: Evaluation of European ancestry-only versus multi-ancestry type 2 diabetes (T2D) polygenic risk scores among childhood cancer survivors in SJLIFE and CCSS

T2D PRS type	Sample	N	OR per PRS SD (95% CI)	P
European-only (Khera <i>et al.</i>)	SJLIFE EUR	3102	2.54 (1.92-3.35)	4.5×10^{-11}
	SJLIFE AFR	574	0.97 (0.39-2.41)	0.95
	CCSS EUR, original	3902	2.16 (1.63-2.88)	1.3×10^{-7}
	CCSS EUR, expansion	2063	3.06 (1.60-5.83)	6.8×10^{-4}
Multi-ancestry (Mahajan <i>et al.</i>)	SJLIFE EUR	3102	1.84 (1.59-2.13)	1.1×10^{-16}
	SJLIFE AFR	574	1.80 (1.22-2.64)	2.8×10^{-3}
	CCSS EUR, original	3902	1.57 (1.36-1.81)	7.8×10^{-10}
	CCSS EUR, expansion	2063	1.74 (1.25-2.42)	9.5×10^{-4}

Abbreviations: SJLIFE, St. Jude Lifetime Cohort; CCSS, Childhood Cancer Survivor Study; EUR, European genetic ancestry; AFR, African genetic ancestry; PRS, polygenic risk score; RT, radiation therapy; OR, odds ratio; SD, standard deviation; CI, confidence interval; P, p-value.

Table 4: *XNDC1N* pathogenic/likely pathogenic rare variation enrichment association with diabetes mellitus risk among childhood cancer survivors in SJLIFE

Rare variant association test	Discovery, SJLIFE EUR (N=1550)			Replication, SJLIFE EUR (N=1552)			Combined, SJLIFE EUR (N=3102)		
	MAF (MAC)	OR (95% CI)	P	MAF (MAC)	OR (95% CI)	P	MAF (MAC)	OR (95% CI)	P
Burden test, <i>XNDC1N</i>	0.004 (13)	-	8.5x10 ⁻⁴	0.003 (8)	-	9.4x10 ⁻³	0.003 (21)	-	8.1x10 ⁻⁶
Collapsed: ≥1 pathogenic variant in <i>XNDC1N</i>	0.004 (13)	8.80 (2.41-32.11)	9.9x10 ⁻⁴	0.003 (8)	12.50 (1.99-78.65)	7.1x10 ⁻³	0.003 (21)	8.65 (3.02-24.74)	5.7x10 ⁻⁵

Abbreviations: EUR, European genetic ancestry; MAF(C), minor allele frequency (count) across pathogenic/likely pathogenic variants mapped to gene; OR, odds ratio; CI, confidence interval; P, p-value.

FIGURES

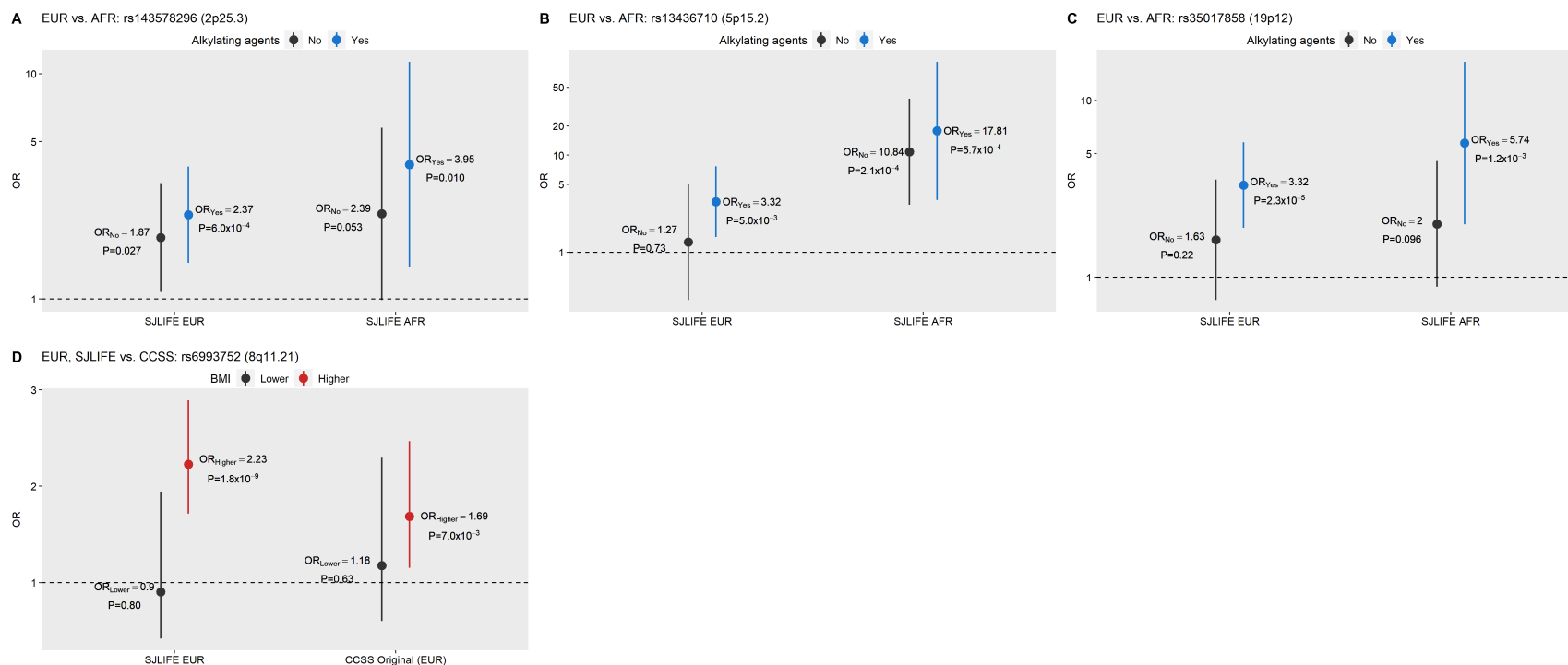


Figure 1: Diabetes mellitus risk associations for replicated index SNPs, stratified by previous treatment with alkylating agents and adiposity. Each panel shows stratified odds ratios (circles) and 95% confidence intervals (whiskers) for the study samples described on the x-axis. Panels A-C show diabetes mellitus risk associations for the risk alleles at the 2p25.3, 5p15.2, and 19p12 loci, respectively, stratified by treatment with alkylating agents (unexposed: black; exposed: blue) in SJLIFE ancestry samples (EUR, European ancestry; AFR, African ancestry). For panels A and C, alkylating agent doses $\geq 4,000$ mg/m² versus none are contrasted, while for panel B, any versus no exposure to alkylating agents are shown (due to lower effect allele frequencies). Panel D shows associations between the risk allele at the 8q11.21 locus (rs6993752) in EUR survivors in SJLIFE and CCSS, stratified by lower body mass index (BMI ≤ 25 kg/m², in black) and higher BMI (> 25 kg/m², highlighted in red).

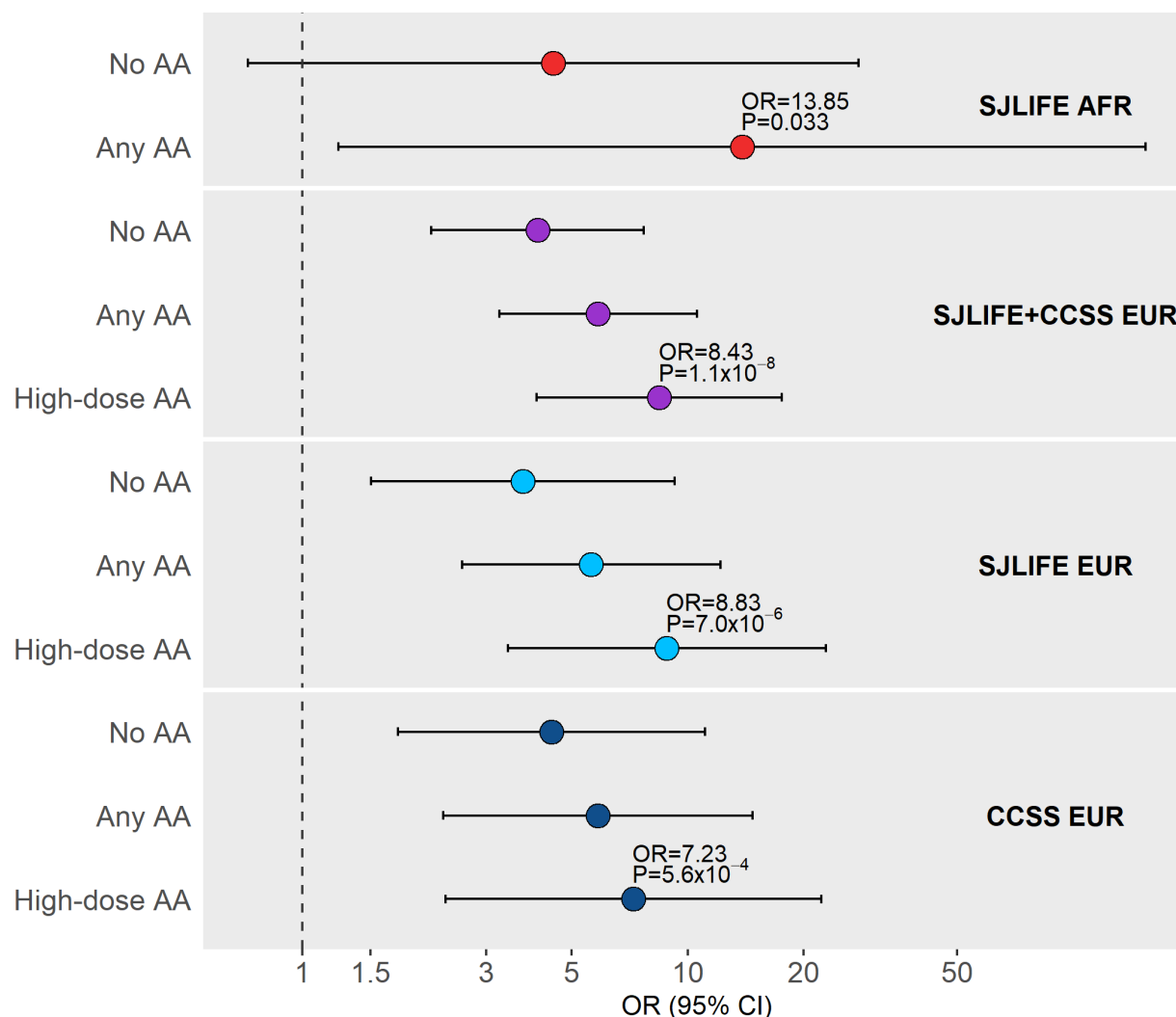


Figure 2: Multi-ancestry PRS quintile associations with diabetes mellitus odds in childhood cancer survivor cohort subgroups defined by ancestry and treatment with alkylating agents (AA). Associations between the top quintile (relative to the lowest quintile) of a general population multi-ancestry PRS for type 2 diabetes (Mahajan *et al.*) and odds of diabetes mellitus in SJLIFE AFR survivors (red), the combined SJLIFE and CCSS cohort of EUR survivors (purple), SJLIFE EUR survivors (light blue), and the combined CCSS EUR survivor cohorts (dark blue) are shown (odds ratios or ORs as circles and corresponding 95% confidence intervals as whiskers). Top PRS quintile ORs and corresponding p-values are annotated for SJLIFE AFR survivors exposed to any AA and for EUR survivors not exposed to abdominal radiation therapy who were treated with high-dose AA (≥ 4000 mg/m²).

METHODS

Study population

For this study, we used data from childhood cancer survivors enrolled in two large retrospectively constructed cohort studies with prospective follow-up of late outcomes: the St. Jude Lifetime Cohort³⁷ (SJLIFE), for discovery analyses, trans-ancestral replication, and trans-ancestry meta-analysis, and the Childhood Cancer Survivor Study³⁸ (CCSS), for EUR-specific GWAS replication and secondary replication of findings from trans-ancestry meta-analysis. This study was approved by the institutional review boards at all participating study centers, including St. Jude Children's Research Hospital (SJCRH). All participants provided written informed consent.

St. Jude Lifetime Cohort

Details of the SJLIFE study methodology have been described previously³⁷. Briefly, SJLIFE includes five-year childhood cancer survivors who were treated for pediatric cancer at SJCRH between 1962 and 2012. All participants included in this investigation underwent at least one on-campus comprehensive medical evaluation and provided a biospecimen for DNA sequencing.

Childhood Cancer Survivor Study

Study design and methods for CCSS have been published in detail elsewhere^{39,40}. Participants include five-year childhood cancer survivors diagnosed at <21 years of age between 1970 and 1986 (described as the CCSS Original Cohort) and between 1987 and 1999 (described as the CCSS Expansion Cohort) who received treatment for pediatric cancer at one of 31 participating study institutions in the United States and Canada. In this analysis, we included all participants who completed a CCSS questionnaire which asked about the presence of chronic health conditions and provided a biospecimen for DNA genotyping or sequencing.

Clinical risk factors and DM ascertainment

Data related to cancer treatments were abstracted from the medical record. Alkylating agent cumulative dose within five years of primary cancer diagnosis was abstracted and quantified as cyclophosphamide-equivalent dose⁴¹. Using data from the review of individual RT records, abdominal RT dose was estimated from the maximum tumor dose or total delivered dose from all overlapping abdominal RT fields⁴². Body mass index (BMI) for SJLIFE participants was calculated from height and weight measurements taken during the most recent campus study visit and adjusted for amputation, while BMI for CCSS participants was calculated from self-reported height and weight measurements from the most recently completed survey.

In SJLIFE, diabetes mellitus (DM) status was clinically ascertained. Participants with symptomatic diabetes, use of oral medications or insulin to treat diabetes, or blood tests consistent with DM diagnosis (fasting blood glucose ≥ 126 mg/dL, random glucose ≥ 200 mg/dL or HbA1c $\geq 6.5\%$) were considered to be DM cases in accordance with the NCI Common Terminology Criteria for Adverse Events (CTCAE) v4.03 classification system⁴³ (CTCAE grades ≥ 2). While the prevalence of type 1 diabetes is low in survivors^{5,6}, further steps were taken to reduce the probability of including DM cases that were unlikely to be T2D. Clinical information was used to exclude participants with transient treatment-related hyperglycemia requiring intervention, as well as participants reporting insulin use and a DM diagnosis before or within 5 years of childhood cancer diagnosis.

In CCSS, we used case definitions of DM consistent with the literature^{5,6} based on self-reported responses from all available baseline and follow-up questionnaires. Survivors who self-reported ever using an oral diabetes medication or insulin to treat diabetes were considered to be cases. Given the potential for misclassification, we also evaluated cases defined by self-report of ever using insulin to treat diabetes in sensitivity analyses.

Genotype data

Methods used to generate genotype data in SJLIFE and CCSS have been described extensively elsewhere⁴⁴⁻⁴⁸. In brief, paired-end WGS was performed using Illumina HiSeq X10 or NovaSeq sequencers for all participants in SJLIFE and the CCSS Expansion Cohort, generating data for common/rare variant association analyses. For the CCSS Original Cohort, DNA was genotyped using the Illumina HumanOmni5Exome array and imputed with the Haplotype Reference Consortium r1.1 reference panel using Minimac3⁴⁹; these data were used for analyses of common variants only. Sample and variant quality control measures were performed separately for the CCSS and SJLIFE autosomal variant data (Supplemental Methods). For each cohort, we identified participants of either European (EUR) or African (AFR) genetic ancestry using EIGENSTRAT-based principal components analyses of independent common variants implemented in PLINK⁵⁰ version 1.9 (Supplemental Methods).

Discovery of common trans-ancestral DM risk loci in SJLIFE

We first conducted ancestry-specific discovery GWAS for clinically-ascertained DM risk in SJLIFE, separately analyzing 3,102 EUR and 574 AFR survivors with complete phenotype and clinical covariate data. In each ancestral sample, common variants were evaluated (minor allele frequency or MAF $\geq 1\%$), including up to 10.5 million variants after excluding variants with excess missingness ($>5\%$) and departures from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$ among controls). Similar to previous GWAS¹⁵, logistic regression models of DM prevalence adjusted for sex, age at last contact, BMI, age at primary cancer diagnosis, cumulative doses of alkylating agents and abdominal RT, and the first ten ancestry-specific principal components were evaluated, assuming an additive genetic effect model.

To facilitate discovery of common loci with multi-ancestral DM risk effects in survivors, we conducted trans-ancestral replication in SJLIFE: top risk associations identified in AFR survivors were subsequently evaluated in EUR survivors and vice versa, using the same models applied in the discovery ancestry-specific GWAS. We prioritized variants from ancestry-specific

discovery analyses with association test p-values $<5 \times 10^{-6}$ that were common ($\text{MAF} \geq 1\%$) across ancestral samples within a 100-kb window of the index variant (i.e., smallest p-value). Common variants meeting a p-value threshold accounting for Bonferroni-corrected multiple testing ($0.05/\text{number of tested variants in the locus with the index variant identified in discovery analysis}$) were considered to be replicated. In addition, trans-ancestry meta-analysis using the fixed-effects inverse variance-weighted method implemented in METAL⁵¹ was performed for index variants. Heterogeneity across genetic ancestral groups was examined using Cochran's Q test⁵² and I^2 index⁵³. Meta-analysis for variants with evidence of ancestral allelic heterogeneity ($P_{\text{het}} < 0.05$) was performed using the Han-Eskin random-effects model (RE2) in METASOFT⁵⁴.

Replication of common DM risk loci in CCSS

Replication was conducted in CCSS EUR survivors (excluding participants also enrolled in SJLIFE) for loci achieving genome-wide significance ($P < 5 \times 10^{-8}$) in EUR-specific GWAS or the trans-ancestry meta-analysis, using the same models applied in SJLIFE. Separate replication analyses were conducted in the CCSS Original ($N=3,902$) and Expansion ($N=2,063$) cohorts due to differences in the derivation of genotype data. If the index variant was unavailable for evaluation in CCSS, we considered the next best variant proxy with the highest estimate of linkage disequilibrium (LD) with the index variant in the 1000 Genomes reference population. Replication was defined by association test p-values < 0.05 in either the CCSS Original or Expansion cohorts.

Treatment- and adiposity-stratified analyses of common DM risk loci

Similar to previous GWAS of late effects in survivors⁵⁵, we evaluated treatment- or adiposity-stratified genetic associations with DM risk exclusively among genome-wide significant index variants with replication evidence to identify whether selected risk factors potentially modified risks for DM in survivors. We adopted this strategy to control the type I error probability

while maximizing power for discovery in small survivor cohorts instead of performing genome-wide scans of gene-treatment interactions, which suffer losses in statistical power with the burden of multiple testing especially when considering several cancer treatment classes and dose definitions. Therefore, we examined replicated index variant associations with DM stratified by the two types of cancer treatments linked to DM^{5,6}: exposures to abdominal RT (none versus any) and alkylating agents (none versus any and lower versus higher dose [$\geq 4,000$ mg/m²] among those with no abdominal RT exposures). Due to the limited sample size of SJLIFE AFR survivors, alkylating agent-stratified analyses were not further stratified by abdominal RT exposures in this sample. Notably, obesity is strongly associated with treatment exposures and DM risk⁶, but abdominal RT is correlated with lower adiposity among survivors¹⁷. Because some DM-associated loci in survivors may not modify risks linked to abdominal RT or alkylating agents but instead modify obesity-associated risks for DM (e.g., involving cranial irradiation), we also conducted adiposity-stratified analyses (BMI ≤ 25 kg/m² versus being overweight or obese, BMI > 25 kg/m²) for loci showing no clear DM risk effect heterogeneity with exposures to these two treatments. All statistical models assessed were adjusted for the same covariates, only omitting specific covariates directly related to the stratification type as appropriate.

External T2D polygenic risk scores (PRSs)

For each survivor, we computed two published T2D PRSs derived from GWAS performed in non-cancer populations. First, we calculated a ~6.9 million variant “European ancestry-only” PRS, derived from a European-ancestry T2D GWAS with 26,676 cases and 132,532 controls⁵⁶, which was subsequently developed and validated in the predominantly European UK Biobank cohort¹⁸ (PGS Catalog: PGS000014). We also calculated a “multi-ancestry” PRS with 338 variants from a GWAS meta-analysis with 180,834 cases and >1.1 million controls (~49% non-European ancestry)¹¹. We implemented recommended best

practices for PRS computation⁵⁷ to calculate all PRSs using original publication weights to support comparability across studies, including common biallelic risk variants (minor allele frequency $\geq 1\%$) of unambiguous strand. All PRSs were generated in our data using PLINK v1.90b (--score)⁵⁰. Adjusted associations between DM risk and standard deviation increases in scaled PRS values and PRS quintiles were evaluated among SJLIFE and CCSS survivors.

Genome-wide analysis of genes enriched with pathogenic rare variant burden

For genome-wide gene-based rare variant burden analyses, we randomly split the SJLIFE EUR cohort *a priori* into approximately equal-sized discovery (N=1,550) and replication (N=1,552) datasets using the median last contact follow-up date (October 15, 2019). Given our limited sample size for discovery, we conducted gene-based rare variant burden tests⁵⁸⁻⁶¹, an approach that compares the number of individuals carrying specific types of rare variants at a given gene among cases and controls and which is most powerful when the set of rare variants affect risk in the same direction and with similar magnitude⁶², using EPACTS⁶³. Specifically, we considered rare variants (MAF<1% and minor allele counts ≥ 3) that were annotated as pathogenic/likely pathogenic (P/LP) based on functional/deleteriousness prediction algorithms, aggregated by gene with respect to the Ensembl⁶⁴ (release 105) gene model (see Supplementary Methods). Associations with DM risk for genes enriched with protein-altering variant burden (with ≥ 2 variants per gene) with discovery association test p-values $< 1.0 \times 10^{-3}$ were prioritized and were considered to be replicated if the gene-based P/LP burden met a p-value threshold accounting for multiple comparisons (0.05/number of prioritized genes from the discovery analysis) in the replication set. Among genes with replicated rare variant burden test associations, the p-value threshold for genome-wide significance for the combined study sample was set at $P < 7.4 \times 10^{-5}$, accounting for a total of 679 genes with protein-altering variant burden enrichments.

Bioinformatics analyses of genome-wide significant DM risk loci

We examined genome-wide significant index variants that were successfully replicated for associations with diabetes mellitus risk in the NHGRI-EBI GWAS Catalog⁶⁵. Index variant associations in the general population using publicly available GWAS summary statistics for EUR participants in the UK Biobank Study⁶⁶ for both self-reported T2D (N=361,141, with 2,292 cases) and ICD-10-coded T2D (N=361,194, with 888 cases) were extracted (<http://www.nealelab.is/uk-biobank/>). *In silico* functional/regulatory annotations for all genome-wide significant DM risk loci were obtained using external genomic data resources (Supplemental Methods).

Data availability

Genotype and phenotype data for all SJLIFE survivors and a subset of CCSS survivors whose childhood cancer was diagnosed between 1987 and 1999 are accessible through the St. Jude Cloud (<https://stjude.cloud>). For CCSS survivors diagnosed between 1970 and 1986, genotype data are available through the database of Genotypes and Phenotypes (dbGaP accession number: phs001327.v2.p1) and relevant phenotype data may be requested through <https://ccss.stjude.org/>.

Acknowledgements

This study was supported by the US National Cancer Institute (R21 CA261833, C Im/Y Yuan principal investigators; R01 CA216354, Y Yasui/J Zhang principal investigators; U01 CA195547, MM Hudson/K Ness, principal investigators; R01 CA261898, P Burrige/Y Sapkota, principal investigators; U24 CA55727, GT Armstrong, principal investigator; Cancer Center Support [CORE] Grant CA21765, C Roberts, principal investigator), and the American Lebanese Syrian Associated Charities.

REFERENCES

1. Howlader, N., *et al.* SEER cancer statistics review, 1975–2018. *National Cancer Institute* (2021).
2. Oeffinger, K.C., *et al.* Chronic health conditions in adult survivors of childhood cancer. *N. Engl. J. Med.* **355**, 1572-1582 (2006).
3. Bhakta, N., *et al.* The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *Lancet* **390**, 2569-2582 (2017).
4. Phillips, S.M., *et al.* Survivors of childhood cancer in the United States: prevalence and burden of morbidity. *Cancer Epidemiology and Prevention Biomarkers* **24**, 653-663 (2015).
5. Mostoufi-Moab, S., *et al.* Endocrine abnormalities in aging survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *J. Clin. Oncol.* **34**, 3240 (2016).
6. Meacham, L.R., *et al.* Diabetes mellitus in long-term survivors of childhood cancer: increased risk associated with radiation therapy: a report for the childhood cancer survivor study. *Arch. Intern. Med.* **169**, 1381-1388 (2009).
7. Friedman, D.N., *et al.* Radiation dose and volume to the pancreas and subsequent risk of diabetes mellitus: a report from the childhood cancer survivor study. *JNCI: Journal of the National Cancer Institute* **112**, 525-532 (2020).
8. De Vathaire, F., *et al.* Radiation dose to the pancreas and risk of diabetes mellitus in childhood cancer survivors: a retrospective cohort study. *The lancet oncology* **13**, 1002-1010 (2012).
9. Friedman, D.N., *et al.* Radiation Dose and Volume to the Pancreas and Subsequent Risk of Diabetes Mellitus: A Report from the Childhood Cancer Survivor Study. *J. Natl. Cancer Inst.* **112**, 525-532 (2020).
10. Vujkovic, M., *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680-691 (2020).
11. Mahajan, A., *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54**, 560-572 (2022).
12. Xue, A., *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 1-14 (2018).
13. Dornbos, P., *et al.* A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. *Nat. Genet.* **54**, 1609-1614 (2022).
14. Im, C., *et al.* Generalizability of "GWAS hits" in clinical populations: Lessons from childhood cancer survivors. *Am. J. Hum. Genet.* **107**, 636-653 (2020).
15. Richard, M.A., *et al.* Germline Genetic and Treatment-Related Risk Factors for Diabetes Mellitus in Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study and St Jude Lifetime Cohorts. *JCO Precision Oncology* **6**, e2200239 (2022).
16. Martin, A.R., *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584 (2019).
17. Wilson, C.L., *et al.* Genetic and clinical factors associated with obesity among adult survivors of childhood cancer: A report from the St. Jude Lifetime Cohort. *Cancer* **121**, 2262-2270 (2015).
18. Khera, A.V., *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219 (2018).
19. Bonder, M.J., *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49**, 131 (2017).

20. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
21. Fu, D., Calvo, J.A. & Samson, L.D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nature Reviews Cancer* **12**, 104-120 (2012).
22. Fry, R.C., *et al.* Genomic predictors of interindividual differences in response to DNA damaging agents. *Genes Dev.* **22**, 2621-2626 (2008).
23. Dhawan, S., Dirice, E., Kulkarni, R.N. & Bhushan, A. Inhibition of TGF- β signaling promotes human pancreatic β -cell replication. *Diabetes* **65**, 1208-1218 (2016).
24. Doco-Fenzy, M., *et al.* Early-onset obesity and paternal 2pter deletion encompassing the ACP1, TMEM18, and MYT1L genes. *Europ. J. Hum. Genet.* **22**, 471-479 (2014).
25. De Rocker, N., *et al.* Refinement of the critical 2p25. 3 deletion region: the role of MYT1L in intellectual disability and obesity. *Genet. Med.* **17**, 460-466 (2015).
26. Sakae, T.a., *et al.* A Japanese patient with a 2p25. 3 terminal deletion presented with early-onset obesity, intellectual disability and diabetes mellitus: A case report. *Journal of Diabetes Investigation* **13**, 391-396 (2022).
27. Ahola-Olli, A.V., *et al.* Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *The American Journal of Human Genetics* **100**, 40-50 (2017).
28. Spracklen, C.N., *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240-245 (2020).
29. Sakae, S., *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415-1424 (2021).
30. Caldecott, K.W. XRCC1 protein; Form and function. *DNA Repair* **81**, 102664 (2019).
31. Marintchev, A., *et al.* Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. *Nat. Struct. Biol.* **6**, 884-893 (1999).
32. Cuneo, M.J. & London, R.E. Oxidation state of the XRCC1 N-terminal domain regulates DNA polymerase β binding affinity. *Proc Natl Acad Sci USA* **107**, 6805-6810 (2010).
33. Thompson, L.H., Brookman, K.W., Jones, N.J., Allen, S.A. & Carrano, A.V. Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Mol. Cell. Biol.* **10**, 6160-6171 (1990).
34. Shen, M.R., Mohrenweiser, H., Thompson, L.H., Thelen, M.P. & Zdzienicka, M.Z. Mutations in hamster single-strand break repair gene XRCC1 causing defective DNA repair. *Nucleic Acids Res.* **26**, 1032-1037 (1998).
35. Seibold, P., *et al.* XRCC1 polymorphism associated with late toxicity after radiation therapy in breast cancer patients. *International Journal of Radiation Oncology* Biology* Physics* **92**, 1084-1092 (2015).
36. Langsenlehner, T., *et al.* Association between single nucleotide polymorphisms in the gene for XRCC1 and radiation-induced late toxicity in prostate cancer patients. *Radiother. Oncol.* **98**, 387-393 (2011).
37. Hudson, M.M., *et al.* Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. *Pediatr. Blood Cancer* **56**, 825-836 (2011).
38. 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 (2009).
39. Robison, L.L., *et al.* Study design and cohort characteristics of the Childhood Cancer Survivor Study: a multi-institutional collaborative project. *Med. Pediatr. Oncol.* **38**, 229-239 (2002).
40. Leisenring, W.M., *et al.* Pediatric cancer survivorship research: experience of the Childhood Cancer Survivor Study. *J. Clin. Oncol.* **27**, 2319 (2009).

41. Green, D.M., *et al.* The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: a report from the Childhood Cancer Survivor Study. *Pediatr. Blood Cancer* **61**, 53-67 (2014).
42. Howell, R.M., Smith, S.A., Weathers, R.E., Kry, S.F. & Stovall, M. Adaptations to a Generalized Radiation Dose Reconstruction Methodology for Use in Epidemiologic Studies: An Update from the MD Anderson Late Effect Group. *Radiat. Res.* **192**, 169-188 (2019).
43. Hudson, M.M., *et al.* Approach for Classification and Severity-grading of Long-term and Late-onset Health Events among Childhood Cancer Survivors in the St. Jude Lifetime Cohort. *Cancer Epidemiol. Biomarkers Prev.* **26**, 666-674 (2016).
44. Morton, L.M., *et al.* Genome-wide association study to identify susceptibility loci that modify radiation-related risk for breast cancer after childhood cancer. *J. Natl. Cancer Inst.* **109**, djx058 (2017).
45. Wang, Z., *et al.* Genetic Risk for Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. *J. Clin. Oncol.* **36**, 2078-2087 (2018).
46. Sapkota, Y., *et al.* Whole-Genome Sequencing of Childhood Cancer Survivors Treated with Cranial Radiation Therapy Identifies 5p15. 33 Locus for Stroke: A Report from the St. Jude Lifetime Cohort Study. *Clin. Cancer Res.* **25**, 6700-6708 (2019).
47. Sapkota, Y., *et al.* Genome-wide association study in irradiated childhood cancer survivors identifies HTR2A for subsequent basal cell carcinoma. *J. Invest. Dermatol.* **139**, 2042-2045. e2048 (2019).
48. Sapkota, Y., *et al.* A novel locus predicts spermatogenic recovery among childhood cancer survivors exposed to alkylating agents. *Cancer Res.* **80**, 3755-3764 (2020).
49. Das, S., *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284-1287 (2016).
50. Purcell, S., *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).
51. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
52. Cochran, W.G. The combination of estimates from different experiments. *Biometrics* **10**, 101-129 (1954).
53. Ioannidis, J.P., Patsopoulos, N.A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e841 (2007).
54. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* **88**, 586-598 (2011).
55. Im, C., *et al.* Genome-wide association studies reveal novel locus with sex-/therapy-specific fracture risk effects in childhood cancer survivors. *J. Bone Miner. Res.* **36**, 685-695 (2021).
56. Scott, R.A., *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888-2902 (2017).
57. Choi, S.W., Mak, T.S.-H. & O'Reilly, P.F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759-2772 (2020).
58. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics* **83**, 311-321 (2008).
59. Madsen, B.E. & Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
60. Morris, A.P. & Zeggini, E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* **34**, 188-193 (2010).

61. Price, A.L., *et al.* Pooled association tests for rare variants in exon-resequencing studies. *The American Journal of Human Genetics* **86**, 832-838 (2010).
62. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics* **95**, 5-23 (2014).
63. EPACTS: Efficient and Parallelizable Association Container Toolbox.
64. Howe, K.L., *et al.* Ensembl 2021. *Nucleic acids research* **49**, D884-D891 (2020).
65. Buniello, A., *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005-D1012 (2019).
66. Bycroft, C., *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203 (2018).