Actionable genetic variants and their associations with late effects risks and mortality among long-term survivors of childhood cancer

SJLIFE Working Groups: Genomics/Genetics
CCSS Working Groups: Genetics; Biostatistics/Epidemiology; Subsequent Neoplasm; Chronic Disease

Investigators

Cindy Im	imcindy@umn.edu
Achal Neupane	achal.neupane@stjude.org
Hasibul Hasan	hasan060@umn.edu
Lucie M. Turcotte	turc0023@umn.edu
Eric J. Chow	ericchow@uw.edu
Daniel A. Mulrooney	daniel.mulrooney@stjude.org
Yutaka Yasui	yutaka.yasui@stjude.org
Kirsten K. Ness	kiri.ness@stjude.org
Melissa M. Hudson	melissa.hudson@stjude.org
Gregory T. Armstrong	greg.armstrong@stjude.org
Kim E. Nichols	kim.nichols@stjude.org
Zhaoming Wang	zhaoming.wang@stjude.org
Yadav Sapkota	yadav.sapkota@stjude.org

Background

The availability of whole-exome sequencing (WES) data has increased across multiple pediatric cancer clinical contexts, informing diagnosis, prognosis, and treatment decision-making[1]. In 2021, the American College of Medical Genetics and Genomics (ACMG) recommended reporting a broad range of “actionable” genetic variants that are identifiable with WES, i.e., pathogenic or likely pathogenic (P/LP) rare genetic variants with sufficient scientific evidence of causing different chronic health conditions (CHCs) with established measures for prevention or treatment[2]. A recent assessment in nearly 58,000 Icelanders[3] found carrying ACMG actionable variants was linked to increased risk for death from diseases that may have been caused by these genetic changes, and carrying any ACMG actionable genetic variant was associated with a 3-year reduction in life span overall. Given childhood cancer survivors face significantly higher risks for developing severe or life-threatening CHCs[4-6] and premature mortality[7,8] compared to the general population, providing information about actionable genetic variants could be critical for informing pediatric cancer patients of their personal lifetime risks for experiencing severe or life-threatening health complications after treatment and premature mortality.

Previous studies among survivors of childhood cancer have largely involved evaluations of the prevalence of P/LP genetic variants in genes with strong cancer predisposition evidence among survivors[9] and their associations with risk for subsequent neoplasms (SN)[10] and subsequent cancer-related mortality[11], or P/LP variants in DNA repair genes

and their association with SN risk[12]. The lack of a global assessment of actionable genotypes among genes in the ACMG Secondary Findings List, which spans cardiovascular, metabolic, and other miscellaneous health conditions, among survivors is a significant knowledge gap. We propose to systematically evaluate the full set of ACMG actionable genetic variants for their contributions to treatment-related morbidity and premature mortality risks among childhood cancer survivors using newly available WES data and detailed clinical annotations for nearly 12,500 survivors participating in the Childhood Cancer Survivor Study (CCSS) and the St. Jude Lifetime Cohort (SJLIFE). Conducting comprehensive assessments of ACMG actionable genetic variants is needed in order to inform how incidental findings may be communicated in the future to patients and providers.

Specific aims

Aim 1a: To characterize the prevalence and distribution of documented actionable variants in genes in the most current American College of Medical Genetics and Genomics (ACMG) Secondary Findings list among 5-year survivors of childhood cancer.

Aim 1b: To compare the prevalence of actionable variants in survivors and the general population (e.g., SJLIFE community controls, TOPMed and All of Us controls, gnomAD non-cancer controls) and individuals with non-cancer diseases (e.g., sickle cell disease).

Aim 1c: To assess if the prevalence of specific actionable variants differs in survivors with corresponding late effects compared to survivors without these late effects.

- Because actionable variants are rare, we will also conduct similar analyses after collapsing specific CHCs into broader disease groups[3], e.g., cancer, cardiovascular disease, and other miscellaneous diseases.

Aim 2: To evaluate associations between carrying actionable variants and incident corresponding CHCs, specific causes of death, and overall survival, accounting for demographic and cancer treatment exposures.

Aim 3 (Exploratory): Enhance understanding of the morbidity and mortality risks posed by ACMG actionable genetic variants among survivors with the development of a web-based summary-level genomic data visualization resource in St. Jude Cloud.

Analytic framework

Study population

For this analysis, we will include all SJLIFE and CCSS 5-year survivors with whole-exome sequencing (WES) data. All analyses described in this concept will be conducted separately in each cohort and within ancestry-specific subgroups, followed by meta-analysis.

Outcome variables

- Vital status, including date of death and cause of death; causes of death should be organized to be consistent with major disease groupings described by Jensson *et al.*[3] (cancer, cardiovascular disease, and other miscellaneous diseases)
- SN history, including ICD-O diagnosis code and date of diagnosis for each SN
- CHCs, graded using the modified CTCAE v4.03 grading system[13], will be evaluated individually as appropriate considering conditions with known associations with ACMG actionable genotypes (https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/), along with major disease groupings described by Jensson *et al.*[3] (cancer, cardiovascular disease, and other miscellaneous diseases)

Sociodemographic/clinical variables

- Sex
- Attained age (i.e., date of birth and date of last follow-up or death)
- Primary cancer diagnosis
- Age at primary cancer diagnosis (i.e., date of primary cancer diagnosis)
- Reported race/ethnicity
- Cancer treatment exposures delivered within 5 years of primary cancer diagnosis
    - Any RT (yes/no)
    - Field-specific RT (yes/no) and dose for each of the 7 major body regions (cranial, neck, chest, abdomen, pelvic, arm, leg)
    - Total body irradiation (yes/no and dose)
    - Chemotherapy
        - Any chemotherapy: yes/no
        - Alkylating agents: yes/no and quantified as cyclophosphamide-equivalent dose[14] (CED)
        - Anthracyclines: yes/no and quantified as doxorubicin-equivalent dose[15] (DED)
        - Epipodophyllotoxins: yes/no and dose
        - Platinum: yes/no and dose
        - Methotrexate (by route): yes/no and dose
        - Steroids: yes/no and dose, if available
        - Vinca alkaloids: yes/no and dose, if available
        - Bleomycin: yes/no and dose
        - Cytarabine: yes/no and dose, if available
        - Mercaptopurine: yes/no and dose, if available
        - Antimetabolites (mercaptopurine and thioguanine): yes/no and dose, if available
    - Hematopoietic cell transplantation (yes/no and type)

Genetic data

WES data for rare variants: Exome-based analyses involving participants in SJLIFE and CCSS Expansion cohorts will use WES data with mean ~54X coverage. A total of 178,399 variants have been removed based on low call rates (<90%), 16,349 variants due to deviations from Hardy-Weinberg Equilibrium ($P<1x10^{-15}$), 19,304 variants due to presence in low-complexity regions, and 321 variants based on minor allele count <1. Sample-level quality control involved evaluation of duplicate samples (none), mismatched WES and WGS data or genetically inferred and self-reported sex (91 samples with sex mismatch), low call rates (<90%), or excess heterozygosity (55 samples outside 3 standard deviations from the mean). A total of 7462 samples (4551: SJLIFE; CCSS Expansion: 2911) and 1,796,854 biallelic variants were retained after implementing these quality control measures. WES data in CCSS Original Cohort will be quality controlled consistent with the procedures described above (approved dbGaP application: project ID 36593). Prior to downstream rare variant analyses, common variants (minor allele frequency or MAF>0.01 in a reference population, e.g., gnomAD, and within our own sample) will be excluded.

P/LP variants

Similar to previous analyses[3,16,17], we intend to analyze variants in selected genes with the following three rare variant "masks":

1. Predicted deleterious missense variants: Annotation will be performed using SnpEff[18]. We will use dbNSFP[19] (version 4.1a), which currently employs ~40 *in silico* prediction tools for annotation. Missense variants will be classified as deleterious if >90% of collated annotations (across all tools) predict deleteriousness.
2. Predicted loss-of-function (LOF) variants: We will use the Loss-of-Function Transcript Effect Estimator (LOFTEE; plug-in implemented in the Variant Effect Predictor or VEP[20] (version 108), see https://github.com/konradjk/loftee). This tool uses VEP to annotate the most severe consequence of a given variant for each gene transcript, while LOFTEE annotates high-confidence loss-of-function (LOF) variants, which include frameshift indels, stop-gain variants and splice site disrupting variants. LOF variants flagged by LOFTEE as dubious (e.g., affecting poorly conserved exons and splice variants affecting NAGNAG sites or non-canonical splice regions) will be excluded.
3. Pathogenic or likely pathogenic (P/LP) variants: NCBI ClinVar[21] will be accessed and searched. Similar to previous work[17,22], the most recent ClinVar adjudications from clinical testing laboratories (2015 onwards) for variants without conflicting interpretations will be used, regardless of phenotype reported in ClinVar given that these may be vague or broad; corresponding literature for phenotypes that appear inconsistent will be reviewed manually on a case-by-case basis.

Statistical analyses

Preliminarily, the prevalence and distribution of P/LP variants across the different sets of ACMG genes, both overall and by each major disease group as described by Jensson

et al.[3] (cancer, cardiovascular disease, and other miscellaneous diseases) will be summarized. Specific P/LP distributions that will be evaluated include: (a) characterization of the P/LP variant classes (e.g., LOF, missense, in-frame deletion/insertion, splice variants), (b) previously documented versus undocumented P/LP variants; and (c) by frequency for each gene. These characteristics will also be compared in controls. Overall survival will be estimated using the Kaplan-Meier method where comparison of survival by P/LP variant carrier status will be based on the log-rank test. The cumulative incidence of CHCs of interest and cause-specific mortality accounting for competing risks will be estimated, where differences by P/LP variant carrier status will be assessed using the Gray K-sample test[23].

We will use Cox regression models with age as the time scale[24] to evaluate associations between carrying ACMG actionable genotypes and outcomes of interest, adjusting for sex, age at childhood cancer diagnosis, cancer treatment decade, the first five genetic ancestry principal components, and treatments (to be defined with clinician investigators; will differ by late effects of interest). To estimate CHC and cause-specific mortality hazard ratios, we will use Cox regression models where individuals are removed (i.e., censored) from the at-risk pool at death.

Analyses described above will also conducted in subgroups stratified by sex, genetic ancestry, and cancer treatment exposures, as appropriate. Meta-analysis of results across cohorts/ancestry groups will use the fixed-effects inverse variance-weighted method[25] and allelic heterogeneity will be examined using Cochran's Q test[26] and the I^2 index[27].

Example tables and figures

Table 1: ACMG actionable genotypes detected among childhood cancer survivors and controls

Disease group	Gene	# P/LP variants	Survivors carrying P/LP variants, % (n)	Controls carrying P/LP variants, % (n)	Specific disease	Inheritance

Figure 1 (example taken from Ke *et al.*[22]): Overview of P/LP variant allele counts across sets of ACMG Secondary Findings genes in SJLIFE/CCSS, where contrasting colors reflect allele counts by case status (e.g., died vs. not; SN-affected vs. not)

- Additional panel (inset pie chart example) showing proportions of cases that can be attributed to specific genes
- Additional panel (example not shown) displaying p-values (Manhattan-style plot with -$\log_{10}$[P] on y-axis and genes on x-axis) for differences in allele prevalence case status by gene (univariate testing)

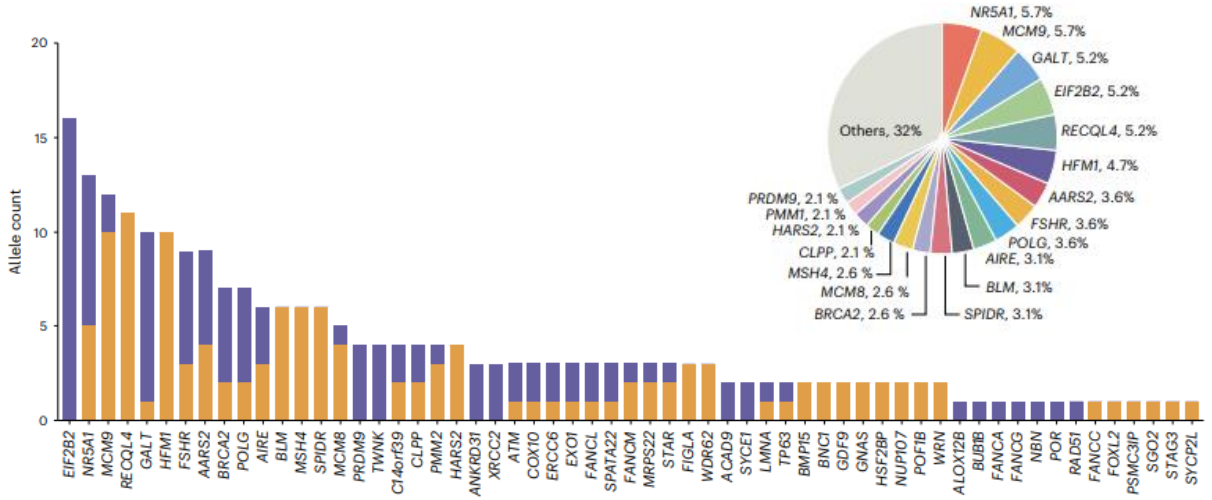

Figure 2: Cumulative incidence of outcomes of interest among carriers and non-carriers of any (or specific sets of) ACMG genotypes in SJLIFE/CCSS

- Overall survival
- Cause-specific mortality
- CHCs by disease group (SNs; cardiovascular diseases; other miscellaneous diseases)
- Additional: stratified by sex, genetic ancestry, cancer treatments

Figure 3: Forest plots showing adjusted hazard ratios and 95% confidence intervals comparing carriers versus non-carriers of any (or specific sets of) ACMG genotypes in SJLIFE/CCSS

- Additional panels showing forest plots from treatment-defined subgroup analyses

References

1. Newman S, Nakitandwe J, Kesserwan CA, et al: Genomes for kids: the scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing. Cancer discovery 11:3008-3027, 2021
2. Miller DT, Lee K, Chung WK, et al: ACMG SF v3. 0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genetics in medicine 23:1381-1390, 2021
3. Jensson BO, Arnadottir GA, Katrinardottir H, et al: Actionable genotypes and their association with life span in Iceland. New England Journal of Medicine 389:1741-1752, 2023
4. Bhakta N, Liu Q, Ness KK, 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
5. Oeffinger KC, Mertens AC, Sklar CA, et al: Chronic health conditions in adult survivors of childhood cancer. N Engl J Med 355:1572-82, 2006
6. Hudson MM, Ness KK, Nolan VG, et al: Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. Pediatric Blood & Cancer 56:825-36, 2011
7. Armstrong GT, Liu Q, Yasui Y, et al: Late mortality among 5-year survivors of childhood cancer: a summary from the Childhood Cancer Survivor Study. J Clin Oncol 27:2328-38, 2009
8. Armstrong GT, Chen Y, Yasui Y, et al: Reduction in late mortality among 5-year survivors of childhood cancer. New England Journal of Medicine 374:833-842, 2016
9. Kim J, Gianferante M, Karyadi DM, et al: Frequency of pathogenic germline variants in cancer-susceptibility genes in the Childhood Cancer Survivor Study. JNCI Cancer Spectrum 5:pkab007, 2021
10. Wang Z, Wilson CL, Easton J, et al: Genetic Risk for Subsequent Neoplasms Among Long-Term Survivors of Childhood Cancer. Journal of Clinical Oncology 36:2078-2087, 2018
11. Chen C, Qin N, Wang M, et al: Cancer germline predisposing variants and late mortality from subsequent malignant neoplasms among long-term childhood cancer survivors: a report from the St Jude Lifetime Cohort and the Childhood Cancer Survivor Study. The Lancet Oncology 24:1147-1156, 2023
12. Qin N, Wang Z, Liu Q, et al: 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. Journal of Clinical Oncology 38:2728, 2020
13. Hudson MM, Ehrhardt MJ, Bhakta N, 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 Epidemiology, Biomarkers & Prevention 26:666-674, 2017
14. Green DM, Nolan VG, Goodman PJ, et al: The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: a report from the Childhood Cancer Survivor Study. Pediatric blood & cancer 61:53-67, 2014
15. Feijen EAM, Leisenring WM, Stratton KL, et al: Derivation of Anthracycline and Anthraquinone Equivalence Ratios to Doxorubicin for Late-Onset Cardiotoxicity. JAMA Oncology 5:864-871, 2019
16. Biddinger KJ, Jurgens SJ, Maamari D, et al: Rare and common genetic variation underlying the risk of hypertrophic cardiomyopathy in a national biobank. JAMA cardiology 7:715-722, 2022
17. Shekari S, Stankovic S, Gardner EJ, et al: Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency. Nature Medicine:1-8, 2023

18. Cingolani P, Platts A, Wang LL, et al: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. Fly 6:80-92, 2012
19. Liu X, Li C, Mou C, et al: dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med 12:103, 2020
20. McLaren W, Gil L, Hunt SE, et al: The Ensembl Variant Effect Predictor. Genome Biol 17:122, 2016
21. Landrum MJ, Chitipiralla S, Brown GR, et al: ClinVar: improvements to accessing data. Nucleic Acids Research 48:D835-D844, 2020
22. Ke H, Tang S, Guo T, et al: Landscape of pathogenic mutations in premature ovarian insufficiency. Nature Medicine 29:483-492, 2023
23. Gray RJ: A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of statistics:1141-1154, 1988
24. Kom EL, Graubard BI, Midthune D: Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. American journal of epidemiology 145:72-80, 1997
25. Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190-2191, 2010
26. Cochran WG: The combination of estimates from different experiments. Biometrics 10:101-129, 1954
27. Ioannidis JP, Patsopoulos NA, Evangelou E: Heterogeneity in meta-analyses of genome-wide association investigations. PloS one 2:e841, 2007