Rare and common variation associated with primary ovarian insufficiency risk in survivors of childhood cancer

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

Investigators

Cindy Im	imcindy@umn.edu
Achal Neupane	achal.neupane@stjude.org
Jessica Baedke	Jessica.baedke@stjude.org
Tomoko Yoshida	tomoko.yoshida@stjude.org
Christine Yu	christine.yu@stjude.org
Lucie M. Turcotte	turc0023@umn.edu
Eric J. Chow	ericchow@uw.edu
Yan Yuan	yyuan@ualberta.ca
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
Sogol Mostoufi-Moab	moab@chop.edu
Zhaoming Wang	zhaoming.wang@stjude.org
Yadav Sapkota	yadav.sapkota@stjude.org
Angela Delaney	angela.delaney@stjude.org

Background

Female survivors of childhood cancer are at elevated risk of cessation of gonadal function before age 40 years, i.e., primary ovarian insufficiency (POI). POI occurs in 1-2% of women in the general population[1,2]; in comparison, an estimated 6% of survivors develop acute ovarian failure (AOF), defined as POI occurring within 5 years of childhood cancer diagnosis, while an additional 9% develop POI beyond this time period after completion of therapy[3]. The most recent estimate of POI prevalence in the Childhood Cancer Survivor Study (CCSS) suggests up to 19% of survivors may experience POI by age 40[4]. Known risk factors for developing POI in this population include radiotherapy directed to the abdominal or pelvic region, total body irradiation, and treatment with alkylating agents[2,5-7]. Recently, risk prediction models for AOF[8] and POI[4] in survivors with comprehensive consideration of treatment predictors have been successfully developed and validated. However, understanding of the contributions of germline genetic variation to POI risk is more limited.

Recent work suggests idiopathic POI in the general population is neither strictly monogenic nor polygenic[9,10]. The heritability of menopausal age is estimated to range from 40-50% in familial studies[11-13]. Genome-wide association study (GWAS) meta-analyses have discovered common variants associated with natural age at menopause in the general population[14-16], with the largest analysis to date by Ruth et al.[16] (~200,000 women) identifying 290 independent genetic signals accounting for up to 38% of the estimated SNP-based heritability. Ruth et al.[16] further reported individuals with the top percentile of values for a polygenic risk score (PRS; ~7 million variants) had a 4.7-fold increased risk of developing POI in independent data (versus 50th percentile, 95% CI: 3.2-7.0), nearly equivalent to the POI risk reported for women with pathogenic/likely pathogenic (P/LP) *FMR1* variants, the leading tested monogenic cause of POI. At the other end of the allelic spectrum, an exome-based analysis by Ke et al.[9] found P/LP

variants among 59 putative POI-causative genes may have accounted for ~19% of cases in their data.

Among childhood cancer survivors, it remains unclear what fraction of POI cases might be attributable to cancer treatments, polygenic or monogenic causes, or the combination of these risk factors. An early GWAS of premature menopause risk in the St. Jude Lifetime Cohort (SJLIFE; 799 survivors, with 30 cases) with array-based genotype data identified a novel risk haplotype at the *NPY2R* locus, where carrying the *NPY2R* haplotype was associated with increased premature menopause risk, especially among survivors exposed to ovarian radiotherapy[17]. Genetic association studies for POI risk have not been conducted in the much larger sample of SJLIFE survivors with sequenced genomes or the sample of CCSS survivors with imputed array-based genotype data, and exome-based analyses of rare variants in SJLIFE and CCSS have not been conducted. To date, there are ~2,500 SJLIFE and ~2,000 CCSS female participants with assessable ovarian status information and genotype data. Interestingly, there is growing evidence supporting an emerging hypothesis that background polygenic risk is a potent modifier of complex disease risks conferred by P/LP disease variants, with several reports showing risks conferred by P/LP variants are multiplicatively greater with increasing background polygenic risk[18-21]. Therefore, using whole-genome/exome sequencing (WGS/WES) and imputed array-based genotype data from childhood cancer survivors in SJLIFE and CCSS, we propose to comprehensively identify and evaluate common and rare POI risk variants and assess their potential to enhance polygenic POI risk prediction.

Specific aims

Aim 1: Characterize the distribution of germline P/LP variants in putative POI-causative genes identified in the literature among long-term female survivors of childhood cancer by ovarian status and exposures to relevant treatment risk factors.

- The prevalence of P/LP variants in POI-causative genes among survivors will also be evaluated against female SJLIFE community controls.

Aim 2: Evaluate whether carrying a P/LP variant in a putative POI-causative gene is associated with increased risk for developing POI, or is modified by background POI risks conferred by common variants (e.g., PRS) and treatment exposures.

Aim 3: Conduct agnostic GWAS and exome-wide association studies to identify novel loci associated with treatment-related POI risk among survivors.

Analytic framework

Study population

- For Aims 1 and 2, we will use the combined cohort of SJLIFE and CCSS 5-year survivors with whole-exome sequencing (WES) data. PRSs will be computed using matched SJLIFE WGS data, CCSS WGS data (cancer diagnosed between 1987-1999) and CCSS imputed array data (cancer diagnosed between 1970-1986).
- For Aim 3, we will use SJLIFE as the discovery cohort and CCSS as the replication cohort.
- All analyses described in this concept will be conducted separately in ancestry-specific subgroups, followed by trans-ancestral replication and meta-analysis.

Exclusions

Participants with the following characteristics will be excluded:

- Survivors with a history of allogeneic blood or bone marrow transplantation
- Survivors treated with cranial RT (CRT) ≥30 Gy, with history of hypothalamic/pituitary region tumors, or suspected multiple pituitary hormone deficiencies due to CRT or due to high risk of hypogonadotropic hypogonadism as the cause of their amenorrhea
- Survivors with AOF likely attributable to known clinical/treatment risk factors, e.g., survivors with high predicted AOF risk (>50%), as defined by treatment-informed risk prediction models by Clark et al.[8] (note that this criteria may instead be defined by a preliminary systematic assessment of treatment risk factors, e.g., high ovarian radiotherapy dose)
- Genetic syndrome (Turner or Down's syndrome)

Outcome variables

SJLIFE

Ovarian status is clinically ascertained in SJLIFE using hormone measurements and medical chart/questionnaire review conducted by an endocrinologist. Age at POI will be based on age at which POI was clinically diagnosed. Specifically, SJLIFE participants ≥16 years at assessment are clinically assigned POI status if they meet either of the following criteria before age 40 years:

1. Lab values consistent with requiring treatment for POI, i.e., follicle stimulating hormone >30 mIU/mL and estradiol <17 pg/mL, and persistent amenorrhea, or
2. Documentation of hormone replacement therapy to treat POI, as indicated by their treating provider/endocrinologist.

CCSS

As described in previous work, baseline and longitudinal follow-up questionnaires querying age at menarche and last menstrual period, current menstrual status, cause of menopause if currently menopausal, pregnancy/childbirth, and use of hormonal contraception will be used to ascertain POI and rule out unrelated causes of absent menses consistent with previous analyses in survivors[4-6,22,23]. Five-year survivors without surgical premature menopause aged ≥18 years at the time of questionnaire will be assigned POI status if they: (a) never menstruated; (b) experienced their last menses within 5 years of cancer diagnosis; or (c) menstruated >5 years after cancer diagnosis but experienced their last menses before age 40. For POI assignment, last menses must be reported >12 months before the questionnaire date. Note that in applying these definitions, AOF cases likely attributable to known clinical/treatment risk factors will be excluded from this analysis. Ambiguous cases will be reviewed by an endocrinologist. Self-reported age at POI or surgical menopause will be used in analyses; for survivors who never menstruated, POI age will be assigned at 16 years.

Sociodemographic/clinical variables

- Sex (female only)
- Attained age (i.e., date of birth and date of last follow-up or death)
- Vital status, including date of death
- Primary cancer diagnosis

- Age at primary cancer diagnosis (i.e., date of primary cancer diagnosis)
- Surgical premature menopause including date of procedure (bilateral oophorectomy or hysterectomy if CCSS)
- Subsequent neoplasm history, including date of diagnosis
- SJLIFE only: hormone measurements (estradiol; follicle stimulating hormone) and date of measurement
- Cancer treatment exposures delivered within 5 years of primary cancer diagnosis
    - Pelvic RT (yes/no and dose)
    - Abdominal RT (yes/no and dose)
    - Ovarian RT (yes/no and dose)
    - Total body irradiation (yes/no and dose)
    - Chemotherapy (yes/no and dose, if available)
        - Alkylating agents
            - Yes/no and quantified as cyclophosphamide-equivalent dose[24] (CED)
            - Specific alkylating agents informing CED (yes/no and dose)
    - 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.

WGS or imputed array data for common variants: Quality-controlled genotypes from joint germline variant calling of whole-genome sequencing (WGS) data using Illumina HiSeq X10 or NovaSeq platforms (30X mean coverage) for SJLIFE survivors and CCSS survivors whose primary cancers were diagnosed between 1987-1999 will be used. For CCSS survivors whose primary cancers were diagnosed before 1987, we will use Illumina HumanOmni5Exome array genotype data imputed with the Haplotype Reference Consortium r1.1 reference panel using Minimac3[25]. Sample and variant quality control measures will be performed separately for the CCSS and SJLIFE autosomal variant data, consistent with previous genetic association analyses[26-30].

P/LP variants

We will evaluate a combined set of 105 genes assessed by Shekari et al.[10] and 95 genes Ke et al.[9] ascribed to be causative for POI. Consistent with these previous analyses, 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[31]. We will use dbNSFP[32] (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[33] (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[34] will be accessed and searched. Similar to previous work[9,10], 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.

PRS

A polygenic risk score (PRS) including germline genetic risk variants identified in the most current, largest general population GWAS meta-analysis of menopausal timing[16] will be evaluated, computed with previously described methods[35,36]. Specifically, we will assess a PRS comprised of ~290 previously identified genome-wide significant ($P<5x10^{-8}$) signals (reported by Ruth et al.[16]), along with a PRS described by Ruth et al.[16], built using published summary statistics and the LDPred[37] methodology.

Statistical analyses

Aim 1: Preliminarily, the prevalence and distribution of P/LP variants across the set of POI-causative genes will be summarized among survivors by ovarian status, including: (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 a control sample (e.g., SJLIFE community controls). Given risks for developing POI and ovarian radiotherapy and alkylating agents are well-established, we intend to perform stratified analyses considering the following major treatment subgroups:

1. No ovarian (or abdominal/pelvic) RT versus any, or lower- versus higher-dose (e.g., <10 Gy versus ≥10 Gy for ovarian RT);
2. No alkylator exposure versus any, and lower versus higher doses of alkylators (e.g., <4,000 mg/m^2 versus ≥4,000 mg/m^2), specifically among survivors who were not treated with ovarian (or abdominal/pelvic) RT; and
3. No/low-dose ovarian (or abdominal/pelvic) RT and no/low-dose alkylator exposure versus treatment with either exposure or both exposures.

Time at risk for POI will start at diagnosis of childhood cancer, ending at POI occurrence or at age 40 years, death, first SMN (including those occurring within the first five years), surgical menopause or last follow-up, whichever occurs first. Cumulative incidence of POI accounting for death or surgical menopause as a competing risk will be estimated for survivors stratified by P/LP variant carrier status. Differences in cumulative incidence by P/LP variant carrier status will be assessed using the Gray K-sample test[38]. Cumulative incidence by P/LP variant carrier

status in treatment-stratified subgroups will also be evaluated using the schema described above.

Aim 2: We will use Cox regression models with age as the time scale[39] to model the effect of carrying POI gene P/LP variants on the POI-specific hazard rate, adjusting for age at childhood cancer diagnosis, the first five genetic ancestry principal components, study cohort, and treatments, including cumulative maximum ovarian or pelvic/abdominal RT dose[40] and alkylating agent dose as cyclophosphamide-equivalent dose[24]. Statistical analyses will be performed first among all survivors, followed by analyses within treatment-stratified subgroups and then within tertiles or quartiles of the standardized menopausal timing PRS (i.e., evaluate P/LP status among survivors stratified by different PRS quantiles, adjusting for all other model covariates). Evaluation of prediction performance metrics using previously described methods[4] will also be undertaken, comparing genetic risk predictors (e.g., PRS and P/LP variant carrier status) to a validated POI clinical risk prediction model[4].

Aim 3: GWAS, including approaches to test gene-treatment interactions on a genome-wide scale[35], will be conducting using published methods described in our previous work[28,29,35,41-43]. Our primary strategy for agnostic rare variant analysis is to conduct: (a) individual testing for variants with allele counts >5 using an additive model, and (b) gene-based rare variant burden testing[44-47], 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 magnitude[48], as implemented in EPACTS[49]. Specifically, we will consider rare variants (MAF<1% and minor allele counts ≥3) annotated as P/LP using the three rare variant masks described previously, aggregated by gene with respect to the Ensembl[50] (release 105) gene model. Associations with POI risk for single variants and genes enriched with protein-altering variant burden (with ≥2 variants per gene) with discovery association test p-values $<1.0 \times 10^{-3}$ (in SJLIFE) will be prioritized for replication analyses (in CCSS) and will be considered to be replicated if the variant or gene-based P/LP burden meets a p-value threshold accounting for multiple comparisons (0.05/number of prioritized variants or genes from the discovery analysis) in the replication data. Among variants or genes with replicated risk associations, the p-value threshold for genome-wide significance for the combined study sample will be set considering the total number of variants tested or the total number of genes with protein-altering variant burden enrichments. Similar to previous work[41], genome-wide significant variants or genes will be evaluated post-hoc for their potential to modify ovarian RT- or alkylator-related POI risk.

Example Tables and Figures

Below, we provide hypothetical tables/figures primarily to describe how results for rare variant association analyses may be presented. Numerous examples of tables/figures displaying agnostic GWAS and GxT GWAS results are available in our published work[28,29,35,41-43], including our diabetes mellitus GWAS manuscript in press at the *Journal of Clinical Oncology* (Im *et al.*, 2024).

Table 1: Clinical and treatment characteristics in CCSS/SJLIFE, in genetic ancestry-specific subgroups

	European (EUR) N (%) or median (IQR)	African (AFR) N (%) or median (IQR)	Admixed American (AMR) N (%) or median (IQR)
Median age at cancer diagnosis, years (IQR)			
Cancer Diagnosis Leukemia Hodgkin disease Kidney tumors Bone cancer Central nervous system tumors Neuroblastoma Non-Hodgkin lymphoma Soft tissue sarcoma Other			
Minimum ovarian radiation dose, Gy None <10 10 to <20 ≥ 20 Missing			
Abdominal radiation dose, Gy None <10 10 to <20 ≥20 Missing			
Pelvic radiation dose, Gy None <10 10 to <20 ≥20 Missing			
Total body radiation dose, Gy None <10 10 to <20 Missing			
Alkylating agent dose (CED, mg/m^2) None <4000 4000 to <8000 ≥8000 Missing			
POI prevalence			

Abbreviations: POI, primary ovarian insufficiency; IQR, interquartile range; CI, 95% confidence interval; Gy, Gray; mg, milligrams; m, meters; CED, cyclophosphamide equivalent dose.

Table 2: Comparison of POI risk prediction models with non-genetic clinical predictors only versus models including a POI polygenic risk predictor in CCSS/SJLIFE

Risk prediction metrics	Clinical risk score (based on PMID: 37972608)	POI gene P/LP variant carrier status	POI PRS
All female survivors		P-value	P-value
SBrS (95% CI)			
AUPRC (95% CI)			
AUROC (95% CI)			
Any ovarian RT and alkylating agent chemotherapy		P-value	P-value
SBrS (95% CI)			
AUPRC (95% CI)			
AUROC (95% CI)			

Abbreviations: CI: confidence interval; SBrS, scaled Brier score; AUROC, area under the ROC curve; AUPRC, area under the precision-recall curve; PRS, polygenic risk score.

Table 3: Top rare variant and gene-based aggregate burden associations with POI risk in SJLIFE and CCSS (note that ancestry-specific analyses will be conducted)

	Discovery, SJLIFE EUR			Replication, CCSS EUR			Meta-analysis		
SNV/indel or gene[a] (genomic location [sites included])	EAF or EAC	OR (95% CI)	P	EAF or EAC	OR (95% CI)	P	EAF or EAC	OR (95% CI)	P

Abbreviations: EUR, European genetic ancestry; EAF(C), effect allele frequency (or total EA count for gene); OR, odds ratio; CI, confidence interval; P, p-value.

a. Location to be indicated as chr:BP:NEA:EA (SNV/indel) or cytogenetic location (genes); sites included refers to the total number of genomic markers included for gene-based aggregate burden associations.

Figure 1 (example taken from Ke *et al.*[9]): Overview of P/LP variant allele counts across known POI genes in SJLIFE/CCSS (contrasting colors in allele count bars: POI cases versus controls)

- Additional panel (inset pie chart example) showing proportions of POI cases that can be attributed to specific POI 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 between cases and controls by gene (univariate testing)

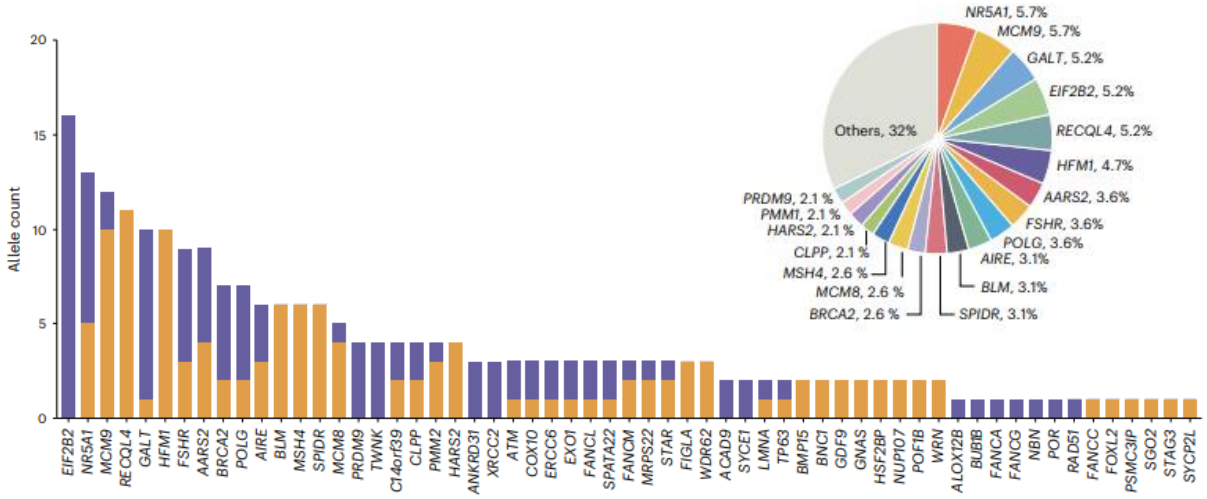

Figure 2: POI cumulative incidence among carriers and non-carriers of POI gene P/LP variants in SJLIFE/CCSS

- Additional panels showing POI cumulative incidence in treatment-defined subgroups:
  - Ovarian (or abdominal/pelvic) RT: None versus any or lower versus higher dose
  - Alkylators: None versus any or lower versus higher dose, specifically among survivors who were not treated with ovarian (or abdominal/pelvic) RT
  - If aforementioned subgroup sample sizes are insufficient: no/low-dose ovarian (or abdominal/pelvic) RT and no/low-dose alkylator exposure versus treatment with either exposure or both exposures

Figure 3: Forest plots showing adjusted hazard ratios and 95% confidence intervals comparing carriers versus non-carriers of POI gene P/LP variants in SJLIFE/CCSS

- Additional panels showing forest plots from treatment-defined subgroup analyses (see above)

Figure 4: P-values for gene-based aggregate burden association result (Manhattan-style plot, $-\log_{10}$P on y-axis and genes on x-axis)

References

1. Mishra GD, Pandeya N, Dobson AJ, et al: Early menarche, nulliparity and the risk for premature and early natural menopause. Human Reproduction 32:679-686, 2017
2. Webber L, Davies M, Anderson R, et al: ESHRE Guideline: management of women with premature ovarian insufficiency. Human Reproduction 31:926-937, 2016
3. Levine JM, Whitton JA, Ginsberg JP, et al: Nonsurgical premature menopause and reproductive implications in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. Cancer 124:1044-1052, 2018
4. Im C, Lu Z, Mostoufi-Moab S, et al: Development and validation of age-specific risk prediction models for primary ovarian insufficiency in long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study and St Jude Lifetime Cohort. The Lancet Oncology, 2023
5. Chemaitilly W, Li Z, Krasin MJ, et al: Premature ovarian insufficiency in childhood cancer survivors: a report from the St. Jude Lifetime Cohort. The Journal of Clinical Endocrinology & Metabolism 102:2242-2250, 2017
6. Mostoufi-Moab S, Seidel K, Leisenring WM, et al: Endocrine abnormalities in aging survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. Journal of Clinical Oncology 34:3240, 2016
7. Johnston RJ, Wallace WHB: Normal ovarian function and assessment of ovarian reserve in the survivor of childhood cancer. Pediatric blood & cancer 53:296-302, 2009
8. Clark RA, Mostoufi-Moab S, Yasui Y, et al: Predicting acute ovarian failure in female survivors of childhood cancer: a cohort study in the Childhood Cancer Survivor Study (CCSS) and the St Jude Lifetime Cohort (SJLIFE). The Lancet Oncology 21:436-445, 2020
9. Ke H, Tang S, Guo T, et al: Landscape of pathogenic mutations in premature ovarian insufficiency. Nature Medicine 29:483-492, 2023
10. Shekari S, Stankovic S, Gardner EJ, et al: Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency. Nature Medicine:1-8, 2023
11. van Asselt KM, Kok HS, Pearson PL, et al: Heritability of menopausal age in mothers and daughters. Fertility and sterility 82:1348-1351, 2004
12. Murabito JM, Yang Q, Fox C, et al: Heritability of age at natural menopause in the Framingham Heart Study. The Journal of Clinical Endocrinology & Metabolism 90:3427-3430, 2005
13. Morris DH, Jones ME, Schoemaker MJ, et al: Familial concordance for age at natural menopause: results from the Breakthrough Generations Study. Menopause 18:956-961, 2011
14. Day FR, Ruth KS, Thompson DJ, et al: Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nature genetics 47:1294-1303, 2015
15. Stolk L, Perry JR, Chasman DI, et al: Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nature genetics 44:260-268, 2012
16. Ruth KS, Day FR, Hussain J, et al: Genetic insights into biological mechanisms governing human ovarian ageing. Nature 596:393-397, 2021
17. Brooke RJ, Im C, Wilson CL, et al: A high-risk haplotype for premature menopause in childhood cancer survivors exposed to gonadotoxic therapy. Journal of the National Cancer Institute 110:895-904, 2018
18. Fahed AC, Wang M, Homburger JR, et al: Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nature communications 11:1-9, 2020

19. Mars N, Widén E, Kerminen S, et al: The role of polygenic risk and susceptibility genes in breast cancer over the course of life. Nature communications 11:1-9, 2020
20. Hassanin E, May P, Aldisi R, et al: Breast and prostate cancer risk: The interplay of polygenic risk, rare pathogenic germline variants, and family history. Genetics in Medicine 24:576-585, 2022
21. 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
22. Chemaitilly W, Mertens AC, Mitby P, et al: Acute ovarian failure in the childhood cancer survivor study. The Journal of Clinical Endocrinology & Metabolism 91:1723-1728, 2006
23. Sklar CA, Mertens AC, Mitby P, et al: Premature menopause in survivors of childhood cancer: a report from the childhood cancer survivor study. Journal of the National Cancer Institute 98:890-896, 2006
24. 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
25. Das S, Forer L, Schönherr S, et al: Next-generation genotype imputation service and methods. Nature genetics 48:1284-1287, 2016
26. Morton LM, Sampson JN, Armstrong GT, et al: Genome-wide association study to identify susceptibility loci that modify radiation-related risk for breast cancer after childhood cancer. Journal of the National Cancer Institute 109:djx058, 2017
27. 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
28. Sapkota Y, Cheung YT, Moon W, 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. Clinical Cancer Research 25:6700-6708, 2019
29. Sapkota Y, Turcotte LM, Ehrhardt MJ, et al: Genome-wide association study in irradiated childhood cancer survivors identifies HTR2A for subsequent basal cell carcinoma. Journal of Investigative Dermatology 139:2042-2045. e8, 2019
30. Sapkota Y, Wilson CL, Zaidi AK, et al: A novel locus predicts spermatogenic recovery among childhood cancer survivors exposed to alkylating agents. Cancer research 80:3755-3764, 2020
31. 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
32. 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
33. McLaren W, Gil L, Hunt SE, et al: The Ensembl Variant Effect Predictor. Genome Biol 17:122, 2016
34. Landrum MJ, Chitipiralla S, Brown GR, et al: ClinVar: improvements to accessing data. Nucleic Acids Research 48:D835-D844, 2020
35. Im C, Yuan Y, Austin ED, et al: Leveraging therapy-specific polygenic risk scores to predict restrictive lung defects in childhood cancer survivors. Cancer research 82:2940-2950, 2022
36. Im C, Sharafeldin N, Yuan Y, et al: Polygenic Risk and Chemotherapy-Related Subsequent Malignancies in Childhood Cancer Survivors: A Childhood Cancer Survivor Study and St Jude Lifetime Cohort Study Report. Journal of Clinical Oncology 41:4381-4393, 2023

37. Vilhjálmsson BJ, Yang J, Finucane HK, et al: Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The american journal of human genetics 97:576-592, 2015
38. Gray RJ: A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of statistics:1141-1154, 1988
39. 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
40. Howell RM, Smith SA, Weathers RE, et al: Adaptations to a Generalized Radiation Dose Reconstruction Methodology for Use in Epidemiologic Studies: An Update from the MD Anderson Late Effect Group. Radiation research 192:169-188, 2019
41. Im C, Li N, Moon W, et al: Genome-wide association studies reveal novel locus with sex-/therapy-specific fracture risk effects in childhood cancer survivors. Journal of Bone and Mineral Research 36:685-695, 2021
42. Sapkota Y, Ehrhardt MJ, Qin N, et al: A Novel Locus on 6p21. 2 for Cancer Treatment–Induced Cardiac Dysfunction Among Childhood Cancer Survivors. Journal Of The National Cancer Institute 114:1109-1116, 2022
43. Sapkota Y, Qin N, Ehrhardt MJ, et al: Genetic Variants Associated with Therapy-Related Cardiomyopathy among Childhood Cancer Survivors of African Ancestry. Cancer Res 81:2556-2565, 2021
44. Li B, Leal SM: 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
45. Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. PLoS genetics 5:e1000384, 2009
46. Morris AP, Zeggini E: An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genetic epidemiology 34:188-193, 2010
47. Price AL, Kryukov GV, de Bakker PI, et al: Pooled association tests for rare variants in exon-resequencing studies. The American Journal of Human Genetics 86:832-838, 2010
48. Lee S, Abecasis GR, Boehnke M, et al: Rare-variant association analysis: study designs and statistical tests. The American Journal of Human Genetics 95:5-23, 2014
49. EPACTS: Efficient and Parallelizable Association Container Toolbox.,
50. Howe KL, Achuthan P, Allen J, et al: Ensembl 2021. Nucleic Acids Research 49:D884-D891, 2020