Childhood Cancer Survivor Study
Proposal for Genome-Wide Investigation of Late-Effects
Late-effects outcome: dyslipidemia

Title: Genome-wide association to study dyslipidemia in adult survivors of childhood cancer

Working group and investigators:
CCSS Chronic Disease Working Group
Drs. Vincent G. Pluimakers, MD, PhD-candidate
Prof. Marry M. van den Heuvel-Eibrink, MD, PhD
Dr. Sebastian J.C.M.M. Neggers, MD, PhD
Prof. André G. Uitterlinden, PhD
Dr. L. Broer, PhD
Dr. G.T. Armstrong, MD, MSCE
Dr. S. Bhatia, MD, MPH
Dr. Y. Yasui, PhD
Dr. K.C. Oeffinger, MD
Dr. L.L. Robison, PhD
Dr. R.M. Howell, PhD
Dr. L.M. Morton, PhD
Dr. J.N. Sampson, PhD
W.M. Leisenring, ScD
Dr. M.M. Hudson, MD
Dr. R.J. Brooke, MD, MSCE, MPH

In the future: collaboration with the DCOG LATER Group

In the future: collaboration with the DCOG LATER Group

1 Childhood Cancer Survivor Study Genetics Working Group Data Access Committee
2 Princess Máxima Centre for Pediatric Oncology, Utrecht, The Netherlands
3 Erasmus Medical Centre, Rotterdam, The Netherlands
4 St. Jude Children’s Research Hospital, Memphis, USA
5 University of Alabama at Birmingham, Birmingham, USA
6 Duke Cancer Institute, Durham, USA
7 The University of Texas MD Anderson Cancer Center, Houston, USA
8 National Cancer Institute, Bethesda, USA
9 Fred Hutchinson Cancer Research Center, Seattle, USA

Correspondence: Vincent Pluimakers, v.g.pluimakers@prinsesmaximacentrum.nl

Background and rationale:
Metabolic syndrome (MetS) is a cluster of adiposity, insulin resistance/diabetes mellitus, dyslipidemia and hypertension. People with MetS are three times more likely to develop a heart attack or stroke and twice as likely to die from cardio- and cerebrovascular disease. In addition, patients with MetS are five times more likely to develop diabetes mellitus type 2 and people with diabetes are three times more likely to develop cardiovascular disease (1-3) (Figure 1). It is estimated that 20-25% of the world’s adult population suffers from MetS (1). Childhood cancer survivors (CCS) are at increased risk of developing (components of) MetS and cardiovascular disease (4, 14-16). Hence, standardized mortality risk due to stroke and coronary heart disease ranged from 1.9 to 12.7 in several epidemiologic studies, with higher risk for specific subgroups with regard to diagnosis, treatment and age (4, 14-16).

For dyslipidemia, the component we propose to study, several large studies in CCS, among which the CCSS, showed increased risk compared to (sibling) controls (16, 17, 18). Our recent literature search identified several therapeutic factors to be associated with dyslipidemia, either directly or through obesity (19). Growth hormone deficiency (GHD), hypothyroidism and hypogonadism can cause dyslipidemia and obesity, hence disease/treatment-related risk factors are brain tumors, brain surgery, cranial radiation and radiation to the thyroid or gonads. Other potential therapeutic factors are abdominal radiotherapy, leading to aberrant lipid storage and metabolism, and corticosteroids.

In addition, there may a be a role for genetic predisposition, as differences in late effects are seen in survivors who received similar cancer therapies (20, 21). As the genetic predisposition of obesity, hypertension and...
diabetes mellitus are already being investigated by other research groups, we propose to study the other denominator of MetS: dyslipidemia.

A genome-wide association study (GWAS) is a non-candidate-driven approach to identify novel genetic variants (mostly single nucleotide polymorphisms (SNPs)) associated with disease. GWAS is typically used to study genetic predisposition in polygenic conditions with genetic variants with small effects. Dyslipidemia is an example of such a condition. Over the past years, numerous loci associated with dyslipidemia have been discovered and replicated in the general population. However, as cohorts of CCS are often small, genetic variation studies for long-term effects in this population are scarce. So far, obesity is the only component of MetS of which the genetic variation in CCS has been studied using GWAS. This study in 1996 survivors of solid and hematological malignancies revealed five SNPs associated with obesity.

Figure 1. The components of the MetS, risk factors for developing the syndrome and the risk for MetS patients to develop cardiovascular disease and type 2 diabetes.

Over the past years, large consortia have performed GWA-studies in the general population on obesity (GIANT and MAGIC), type 2 diabetes (MAGIC, DIAGRAM, and eQTL Gen Consortium) and metabolic syndrome (UK Biobank) and have identified tens of independently associated loci. Four large GWA-studies for dyslipidemia in more than 100,000 individuals (non-childhood cancer survivors) revealed numerous loci that were associated with different lipid traits. The most important findings from these studies are listed in Table 1.
Table 1. Most important findings from large GWA-studies on dyslipidemia in the general population.

<table>
<thead>
<tr>
<th>Article</th>
<th>Population</th>
<th>Variant</th>
<th>Gene</th>
<th>Lipid trait(s)</th>
<th>MAF</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>England 2017 BMC Cancer</td>
<td>209 ALL survivors</td>
<td>rs676210, rs2286615, rs146008363, rs62079523</td>
<td>APOB, BAD, EIF4B, OGFOD3</td>
<td>Low LDL, n.s.</td>
<td>23%, 10%, 5%, 33%</td>
<td>0.0069, 0.00065, 0.00018, 0.00032</td>
</tr>
</tbody>
</table>

Table 2. Results of single study in CCS reporting genetic predisposition to dyslipidemia.

Aim of the study:
To perform the first GWAS on dyslipidemia in childhood cancer survivors and identify novel genes that are associated with dyslipidemia in childhood cancer survivors.

Main outcome:
In the CCSS cohort, dyslipidemia is defined as CTCAE grade 2, based on survivors’ answers to the baseline and baseline expansion health questionnaires[38]. In the SJLIFE cohort, dyslipidemia is defined as the clinical diagnosis based on blood lipids (Table 2)[39].

Study population:
The original cohort of the Childhood Cancer Survivor (CCSS) study is the discovery cohort. This is a cohort of more than 14,000 CCS, diagnosed between 1970 and 1986, and 4000 siblings. All participants in the study completed questionnaires. Of these, 5739 CCS were genotyped. In the CCSS, 48.5% of participants are male and 40% were diagnosed between 0 and 4 years old. The most common malignancies were ALL (29.8%), Hodgkin lymphoma (13.3%) and central nervous system tumors (12.0%). Treatment comprised chemotherapy in 74.1% and radiotherapy in 62.0%.

The St. Jude Lifetime Cohort (SJLIFE) will serve as the replication cohort (the data will be accessed by the St. Jude collaborators). This cohort consists of more than 4000 CCS who completed health questionnaires and also underwent clinical assessment for late effects in the After Completion of Therapy clinic. About 3000 CCS underwent whole genome sequencing. A second replication set will be the CCSS Expansion Cohort, diagnosed between 1987 and 1999, of whom 3000 have completed whole genome and whole exome sequencing and are available on the St. Jude Cloud. The CCSS and CCSS Expansion cohorts will exclude survivors also participating in SJLIFE. Survivors that received hematopoietic stem cell transplantation are excluded.

Genetic predisposition to dyslipidemia in these cohorts will be investigated using GWAS-analysis. So far, genotype data in these cohorts have been used to study genetic predisposition to obesity, premature menopause and subsequent breast cancer in CCS[22, 40, 41].
Exploratory variables:
The following variables will be included as covariates (based on literature):
- Patient characteristics:
  o Sex: male, female.
  o Genetic ancestry: European.
  o Age: age at follow-up.
  o BMI (weight/height²) at follow-up.
  o Smoking.
  o Comorbidities: growth hormone deficiency, hypogonadism, diabetes mellitus, hypothyroidism, cardiomyopathy.
- Childhood cancer type.
- Treatment characteristics:
  o Radiotherapy: cranial, abdominal (with subgroup pancreatic), total body. Prescribed dose, or if available dosimetry.
  o Chemotherapy: total cumulative dose anthracyclines (doxorubicin-equivalent dose), alkylating agents (cyclophosphamide equivalent dose), asparaginase.
  o Steroid exposure: prednisone, dexamethasone; yes/no.

Statistical analysis plan
Quality control of genotype data in CCSS and WGS data of SJLIFE has already been performed. Imputation has been performed with the CCSS genotype data. The SNP-analysis will be performed with the RVtests software package(42), using logistic regression analysis with outcome dyslipidemia yes/no. The analysis will first be performed correcting for age at follow-up, sex and genetic ancestry using PCA. Then, the aforementioned covariates will be added to the model using forward selection, to study whether they influence the SNP-analysis; if so, they will be kept in the model. When necessary due to small sample size, exact logistic regression will be used. After the logistic regression analysis, we will perform survival analysis on identified hits in order to get clinically interesting effect estimates.

Quality control of the SNP-analysis will be performed with the EasyQC package using standard settings(43). This includes filtering based on MAF (excluded when <0.05) and imputation quality (excluded when <0.3, or more stringent <0.5 due to limited replication possibilities).

A genetic power calculator was used to estimate the relative risk that can be found in this study(44) for several minor allele frequencies (MAF). Assuming a sample size of 5,324, a type I error of 5*10^{-6}, a power of 80% and a case-control ratio of 1:4, the relative risk per high risk allele that can be found for MAF = 0.05, 0.1, 0.25 and 0.50 is 1.5, 1.4, 1.3 and 1.3, respectively. Visualization of the genetic associations will be performed with Functional Mapping and Annotation (FUMA)(45). For the association analysis, a significance threshold of p<5*10^{-8} will be used, and a suggestive threshold of p<5*10^{-6}.

SNPs with a p-value below the suggestive threshold will be further analyzed in an interaction model. We will focus on interactions with smoking as well as cranial (SNP*CRT, with categories no, low dose <20Gy, and high dose ≥20Gy) and abdominal radiotherapy (SNP*Abd RT, with categories no, low dose <15Gy, and high dose ≥15Gy). If this data is available, we can also study the interaction with physical activity and socio-economic status.

For the replication analysis, SNPs with a p-value below the suggestive threshold will be tested in the SJLIFE cohort and in the CCSS Expansion Cohort. The SNPs found in the discovery analysis will be sent to the collaborators for replication.

To annotate the biological function of significant replicated findings, a functional analysis will be performed by FUMA, which implements, among others, resources as eQTL, MAGMA and ENCODE/Roadmap. This will be followed by pathway analysis. The tools for pathway analysis will be determined later as there are several developments in this field.

It will be of interest to determine whether found genetic associations are shared with the general population or are unique in CCS. In order to do this, loci and effect sizes in the two large GWA studies mentioned in the introduction will be compared with loci identified in our study, and vice versa. Our team, through André Uitterlinden, also has access to large GWAS consortia (UK Biobank, CHARGE, DIAGRAM, CARDioGRAM).
**Empty tables and figures:**

Table 1. Demographic and treatment characteristics of the discovery and replication cohorts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CCSS Cases</th>
<th>CCSS Controls</th>
<th>SJLIFE Cases</th>
<th>SJLIFE Controls</th>
<th>CCSS Expansion Cases</th>
<th>CCSS Expansion Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyslipidemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic ancestry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at dyslipidemia diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypogonadism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiotherapy dose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cranial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancreas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total body</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthracyclines (DED)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkylating agents (CED)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparaginase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steroids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prednisone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dexamethasone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stem cell transplantation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. SNPs discovered in CCSS cohort and replicated in SJLIFE and/or CCSS Expansion cohorts.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>BP</th>
<th>Gene</th>
<th>RA</th>
<th>EA</th>
<th>EAF</th>
<th>Rsq</th>
<th>OR</th>
<th>95%CI</th>
<th>P</th>
<th>Rsq</th>
<th>OR</th>
<th>95%CI</th>
<th>P</th>
<th>Rsq</th>
<th>OR</th>
<th>95%CI</th>
<th>P</th>
<th>OR</th>
<th>95%CI</th>
<th>P</th>
</tr>
</thead>
</table>

RA = reference allele; EA = effective allele; EAF = effective allele frequency; Rsq = imputation quality in replication

Figure 1. Manhattan plot of the discovery analysis

Figure 2. QQ-plot

Figure 3. Locus zoom

Figure 4. Forest plot of pooled results from discovery and both replications
Timeline:
Our team is motivated to start this project and can give priority to analyzing the data. We propose a two steps approach, with step one being the GWAS-analysis of dyslipidemia in the Original CCSS (diagnosed 1970-1986), SJLIFE and CCSS Expansion (diagnosed 1987-1999) cohorts. Collection of data and analysis of availability of covariates will be performed in May 2020. The association analysis will be performed in June and July, after which the found SNPs will be sent for replication. Further analysis of significant replicated SNPs will take one month. When this is finished, the manuscript will be written. In step two, the findings of this study can be replicated in the national DCOG LATER cohort (see below) and a meta-analysis of the four cohorts will be performed.

Perspective for further replication:
After the Dutch DCOG LATER study is finished in 2020, genetic data from this cohort can be used for further replication – we submitted an application of interest to the DCOG LATER board to use this data. The DCOG LATER cohort is a national cohort of 6165 CCS treated in The Netherlands between 1963 and 2002, of which childhood cancer and treatment characteristics were collected in a central database[46]. In part 1 of the DCOG LATER study, in 2013, a general health questionnaire was sent to the cohort, which contained questions on past and current health status – including dyslipidemia, as well as weight, height, diabetes and hypertension –, medication use, lifestyle, physical activity and family history of cardiovascular disease, diabetes and cancer. The questionnaire was completed by 3169 survivors. In part 2 of the study, for which accrual is almost finished, the cohort is invited to visit the late effects clinic where, among other things, a short version of the health questionnaire is completed, and blood lipids, blood pressure, glucose and body composition (BMI, waist/hip ratio, DXA-scan) are determined, and DNA (from blood or saliva) is collected for genotyping. The aim is to include 2500-3000 survivors in part 2 of the study. Table 3 provides an overview of dyslipidemia definitions in the study cohorts and the possibility for replication analyses.

<table>
<thead>
<tr>
<th>CCSS &amp; CCSS Expansion</th>
<th>SJLIFE</th>
<th>DCOG LATER* (in 2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCAE high total cholesterol or hypertriglyceridemia grade 2: self-reported high cholesterol or triglycerides and on medication</td>
<td>CTCAE high total cholesterol or hypertriglyceridemia grade 2: total cholesterol &gt;300 mg/dL or triglycerides &gt;300 mg/dL</td>
<td>Clinical diagnosis hypercholesterolemia: cholesterol, triglycerides, HDL, LDL</td>
</tr>
</tbody>
</table>

* Pending approval by the DCOG LATER board.

Table 3. Definition of dyslipidemia in CCSS, SJLIFE and DCOG LATER cohorts.

Funding:
Vincent Pluimakers is supported by core funding of the Princess Máxima Centre for Pediatric Oncology (2019-2021) and by the Foundation KiKa (Children Cancerfree) project number 171/ODAS project number 2014-19 (2017-2019).
References:


43. Heid I, Winkler T, Gorski M, Olden M. Easy QC.

