**BACKGROUND AND SIGNIFICANCE**

Anthracycline chemotherapies (doxorubicin, daunorubicin, epirubicin, idarubicin) form a critical backbone for treatment of children with cancer.\(^1\,^2\) However, the clinical utility of anthracyclines is compromised by dose-dependent cardiotoxicity, manifesting as asymptomatic cardiac disease (CD) (function/structure on imaging studies), progressing to clinically-overt congestive heart failure (CHF).

Childhood cancer survivors are at a 5-15 fold increased risk of anthracycline-associated CD/CHF when compared with the general population\(^3\,^4\,^5\,^6\,^7\,^8\,^9\,^10\,^11\). However, significant inter-patient variability in risk is observed, suggesting the role of genetic susceptibility. Once diagnosed with CHF, the 5 year survival is <50%.\(^12\) The high burden of anthracycline-associated CD/CHF-related morbidity and poor prognosis, coupled with variability in risk, suggests a need to identify patients at highest risk, such that targeted interventions can be instituted.

Extant literature suggests that anthracyclines cause cardiotoxicity by imposing oxidative stress\(^13\); the resulting reactive oxygen species (ROS) damage cardiomyocytes\(^14\). Anthracyclines generate ROS by redox cycling, iron complexation, chaotropic effects in mitochondria, and uncoupling of electron transport chain.\(^15\,^16\,^17\) Anthracyclines can disturb antioxidant defense systems and repair pathways.\(^17\,^18\) Alterations in topoisomerase IIβ by anthracyclines may also be important.\(^19\) Previous studies show associations between single nucleotide polymorphisms (SNPs) and anthracycline-associated CD/CHF, providing credence to the proposed mechanism of cardiotoxicity (Fig. 1) or early detection of anthracycline-associated CD/CHF that could be applied to childhood leukemia/lymphoma survivors at highest risk.

**SPECIFIC AIMS**

**Aim 1: Identify genetic variants associated with a-CD/CHF in childhood cancer survivors**

*Aim 1.1: Identify novel genetic variants associated with anthracycline-associated CD/CHF in childhood cancer survivors*

*Aim 1.2: Comprehensively characterize the role of previously-published genetic variants associated with anthracycline-associated CD/CHF in childhood cancer survivors*

We will use the CCSS GWAS data to both identify novel genetic variants associated with anthracycline-associated CD/CHF in long-term survivors of childhood cancer, as well as replicate previously published genetic
variants, in order to develop a comprehensive list of variants that could then be used for functional studies (Aim 2). Our discovery population will be the CCSS subjects. For our replication population, we will include individuals enrolled in COG-ALTE03N1.

**Aim 2:** Define functional relevance of genetic variants using human-induced pluripotent stem cell-derived cardiomyocytes and human myocardial tissue (Blanco lab).

This application takes advantage of the premise that an individual’s state of health (anthracycline-associated CD/CHF) stems from an interaction between their unique genomic makeup and anthracycline exposure. With the unique resources of the CCSS cohort and COG-ALTE03N1 populations in combination with the laboratory-based expertise, the strong collaborative team (with a proven track record) is uniquely equipped to have a clinically meaningful impact on this field.

**PRELIMINARY DATA**

We have the necessary expertise to successfully execute the proposed study. The PI chairs a COG-wide study (COG-ALTE03N1) open at 119 sites. The study uses a matched case-control design to understand the pathogenesis of a-CD/CHF in childhood cancer survivors. Eligibility, and matching are shown in Table 1.

<table>
<thead>
<tr>
<th>Cases – Eligibility</th>
<th>Matching Criteria</th>
<th>Controls – Eligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dx of cancer at age ≤21y, irrespective of current age</td>
<td>1. Primary cancer diagnosis</td>
<td>1. Dx of cancer at ages 21y, irrespective of current age</td>
</tr>
<tr>
<td>2. No prior history of allogeneic BMT</td>
<td>2. Year of cancer diagnosis (±5y)</td>
<td>2. No prior history of allogeneic BMT</td>
</tr>
<tr>
<td>4. Submission of a blood specimen or saliva</td>
<td>4. Time between cancer diagnosis to study (for controls) to exceed time between cancer diagnosis and CD/CHF diagnosis</td>
<td>4. Submission of a blood specimen or saliva</td>
</tr>
<tr>
<td>5. Written informed consent from patient and/or the patient’s legally authorized guardian</td>
<td>5. Written informed consent from patient and/or the patient’s legally authorized guardian</td>
<td></td>
</tr>
</tbody>
</table>

*CD/CHF uses Am Heart Assoc criteria for cardiac compromise, dyspnea/fatigue/edema/hepatomegaly/rales +/- ejection fraction [EF] ≤40%/ fractional shortening [SF] ≤28%

For each case or control, Therapeutic Summary is prepared by the COG site clinical research assistant (CRA), reviewed by the site PI, and submitted to the coordinating center at UAB. Completion rates: cases (99%); controls (99%). Lifetime anthracycline exposure is calculated as previously described.

Chest radiation is captured as a yes/no variable. A self-administered questionnaire (cases, controls/parents if patient <12y) captures race/ethnicity, parental education, annual household income; patient height/weight, comorbidities and medications. Completion rates: 98%. Peripheral blood (6 ml in EDTA tube) or saliva (30%) samples using Oragene®-DNA kits are obtained at scheduled clinic visits. Genomic DNA is isolated from blood (QIAamp/Qiagen kits), or saliva (DNA Genotek prepIT®-L2P kit). COG sites provide source documents for validation of anthracycline-associated CD/CHF (echo report, signs/symptoms of CHF). Thus far, we have enrolled 195 cases with CD/CHF (~90% with anthracycline exposure) and 585 controls (Table 2). These cases and controls have contributed to the genetic association studies described below.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cases (n=195)</th>
<th>Controls (n=585)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at leukemia/lymphoma diagnosis (years)</td>
<td>Median (range)</td>
<td>7.3 (0-20.7)</td>
<td>7.6 (0-21.1)</td>
</tr>
<tr>
<td>Age at study participation (years)</td>
<td>Median (range)</td>
<td>16.6 (0.4-41)</td>
<td>18.5 (2.0-49)</td>
</tr>
<tr>
<td>Females</td>
<td>107 (55%)</td>
<td>527 (49%)</td>
<td>0.2</td>
</tr>
<tr>
<td>Non-Hisp whites</td>
<td>142 (73%)</td>
<td>462 (79%)</td>
<td>matched</td>
</tr>
<tr>
<td>Length of follow-up (years)</td>
<td>Median (range)</td>
<td>7.0 (0.1-35.1)</td>
<td>11.2 (0.4-40.3)</td>
</tr>
<tr>
<td>Cumulative Anthracycline Exposure (mg/m²)</td>
<td>Median (range)</td>
<td>300 (0-575)</td>
<td>140 (0-1050)</td>
</tr>
</tbody>
</table>

Of note, in each of the examples below, the variants were successfully replicated and their functional relevance examined. We have re-opened the CD/CHF arm (currently 112 COG sites are participating). We plan to enroll approximately 150 additional cases by December, 2019. We project to achieve a total of 350 cases with CD/CHF and 700 matched controls for replication within Aim 1. For functional validation (Aim 2), we will obtain blood and ship to Blanco lab for the functional validation of the identified SNPs.
INNOVATION

This study will take advantage of two large unique resources – CCSS and COG-ALTE03N1 and has a high potential for guiding clinical management as shown here:

Cancer survivors: risk-based targeted surveillance for early detection of anthracycline-associated CD/CHF and risk-based targeted pharmacologic/behavioral intervention strategies to reduce risk.

Pathogenesis of anthracycline-related CD/CHF: insights into intervention, based on the function of genetic variants identified to be pathogenically involved in anthracycline-associated CD/CHF.

Single-gene: Carbonyl reductase catalyzes reduction of parent anthracyclines to cardiotoxic metabolites. We found that SNPs in CBR3 (CBR3 V244M) modified the dose-dependent risk of anthracycline-associated CD/CHF. In patients with CBR3:GG (CBR3 M244), exposure to <250mg/m² anthracyclines increased risk when compared to CBR3:GA/AA genotypes not exposed to anthracyclines (OR=5.5, p=0.003), or exposed to <250mg/m² (OR=3.3, p=0.006). There was no safe dose for patients with CBR3:GG (Fig. 3). Patients exposed to >250mg/m² developed anthracycline-associated CD/CHF, irrespective of CBR3 genotype.

ITMAT/Broad/CARe (IBC) cardiovascular SNP array: We identified a SNP rs2232228 in hyaluronan synthase (HAS3) gene that exerts a modifying effect on anthracycline-associated CD/CHF risk (p=5.3x10⁻⁷). In patients with rs2232228 GG genotype, CD/CHF is infrequent and not dose-related (Fig. 4); in patients receiving >250mg/m² anthracyclines, the rs2232228 AA genotype confers an 8.9-fold (p=0.003) increased a-CD/CHF risk c/w GG. We replicated the gene-environment (GxE) interaction. Functional relevance: Hyaluronan produced by HAS3 reduces ROS-induced cardiac injury. We found lower HAS3 mRNA levels in AA healthy hearts c/w GA/GG hearts; anthracycline-associated CD/CHF risk with AA genotype is likely due to inadequate protection from ROS-mediated injury.

GWAS: SNP rs1786814 on CELF4 gene passed the significance cut-off for GxE interaction. In patients with A allele, anthracycline-associated CD/CHF was infrequent, and not dose-related; in those exposed to >250mg/m² anthracyclines, rs1786814 GG genotype conferred a 10.2-fold (95%CI, 3.8-27.3, P <0.0001) increased risk of anthracycline-associated CD/CHF c/w those with GA/AA genotypes and anthracyline dose ≤250mg/m² (Fig. 5). This SNP was replicated independently. Functional relevance: CELF proteins control splicing of TNNT2 (encodes for cTnT: biomarker of myocardial injury). Coexistence of >1 TNNT2 variants results in a temporally split myofilament response to calcium, causing decreased contractility; as found in idiopathic dilated cardiomyopathy. We found that healthy human hearts with GG rs1786814 genotype were more likely to show coexistence of >1 TNNT2 splicing variants (GG: 90.5% vs. GA/AA: 41.7%; P=0.005), suggesting that the association between anthracyclines and CD/CHF was mediated through expression of abnormally spliced TNNT2 variants.

Pathogenesis of anthracycline-associated CD/CHF: The Blanco lab has applied several approaches to characterize SNPs associated with a-CD/CHF. He has examined expression patterns in informative human tissue samples (heart, liver) and tissue fractions (DNA, RNA, microsomes and mitochondria), and has characterized functional interplays between genetic (SNPs, mitochondrial DNA mutations) and epigenetic factors (DNA methylation, micro RNAs) in human tissues.

APPROACH

Objective: Using the resources offered by the CCSS GWAS data, we aim to determine the pathogenesis of anthracycline-related cardiomyopathy in childhood cancer patients. For this study, we will use the following
approach: i) an agnostic approach (GWAS) to determine associations between novel genetic variants and a-CHF in childhood cancer survivors with a sound replication plan using the COG-ALTE03N1 study; ii) replication of a carefully-curated list of previously-published genetic variants associated with anthracycline-associated CD/CHF; iii) define the functional relevance of the genetic variants.

Requested data elements:

Outcome of interest: The primary outcome of interest is CD/CHF; this will be obtained from the CCSS surveys (Baseline, Follow-up 2003, and Follow-up 2007). We will use previously-described methodology to define CHF. Outcomes graded as severe (grade 3; self-reported CD/CHF plus medications), life-threatening (grade 4; requiring heart transplantation), or fatal (grade 5) with available GWAS data will be included. It is projected that we have genome-wide SNP array data for 220 subjects of European ancestry with CD/CHF as defined above. Of note, when data from the most current follow-up is available (i.e., FU5), there may be additional CD/CHF data to include in this study.

Exposures of interest: the primary exposures are genotypes obtained from Illumina Human Omni5Exome array. For Aim 1.2, We have searched OVID Medline, PubMed, Cochrane Central Register of Controlled Studies, CINAHL Plus, AMED, EMBASE and HuGE Navigator from inception until August 2017. Search terms included anthracycline, doxorubicin, daunorubicin, epirubicin, idarubicin, cardiotoxicity, cardiomyopathy, heart failure, SNP, pharmacogenomics and genetic association. This was supplemented with a manual search of cited references from retrieved articles. We found a total of 29 studies, examining 83 different genes (138 SNPs) linked to the biochemical pathway of anthracycline metabolism, oxidative stress or cardiac function. When we are ready for the analysis, we will re-run the search to supplement this gene/ SNP list.

Covariates: Primary cancer diagnosis; Age at Baseline, Follow-up2, Follow-up4, Follow-up5; Gender; Height and weight at Baseline, Follow-up2, Follow-up4, Follow-up5; Education and household income at Baseline, Follow-up2, and Follow-up4, Follow-up5; Physical activity measures when obtained; Radiation therapy to the chest (y/n and dose); Radiation to the heart (yes/no and dose); Anthracyclines (cumulative dose); Other Chemotherapy (y/n); Smoking (ever, never); Diabetes (y/n, age at diagnosis); Hypertension (y/n, age at diagnosis); Dyslipidemia (y/n, age at diagnosis)

Discovery Population: We will leverage genome-wide SNP array data available on childhood cancer survivors enrolled in the CCSS Original Cohort. We anticipate 220 patients with CD/CHF and 5,096 without, based on CCSS Public Access GWAS Data Tables (https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/chronic-gwas.pdf).

Technical Plans for Replication: We will use the resources offered by COG-ALTE03N1 (Chair, S Bhatia) to i) replicate novel genetic variants identified (Aim 1.1: phase I of COG-ALTE03N1: 195 cases); ii) provide material toward functional relevance of identified genetic variants (using fresh blood for iPSC). For the replication, we will ensure that cases and controls that overlap with the CCSS cohort are excluded from the Replication analysis. Our most recent review of the degree of overlap revealed that <10% of the cases overlapped between the two studies. For genotyping the candidate SNPs, SNPType assays (Fluidigm Corporation, South San Francisco, CA) will be designed using the D3 Assay design tool at UAB. Genomic DNA (10-100 ng) will be subjected to genotyping (PCR on Juno 96.96 Genotyping IFC and Juno platform). The IFC will then be transferred to the Biomark™ HD system for end point reads. Genotype calls will be obtained using Fluidigm SNP Genotyping Analysis software.

Statistical Consideration

Quality Control of Genotype data: Stringent quality control checks of the called genotypes will be performed in PLINK 2.0. We will remove SNPs with a call rate <95%, a Hardy-Weinberg equilibrium P <5×10⁻⁵, and a minor allele frequency (MAF) <0.5%. We have purposefully restricted our analyses to MAF <0.5% because, we believe that we will be underpowered to detect rarer variants, given the sample size. To avoid samples who will poor quality DNA, we will drop individuals with sample call rates <95% or heterozygosity rates >±2.5 SD from the array median. For each individual, X chromosome heterozygosity will be evaluated to estimate sex, and individuals with discordant self-reported and genetically-defined sex will be flagged for possible removal. We will estimate cryptic relatedness between individuals using the identical by descent (IBD) procedure. Samples with a pi-hat value >0.2 will be flagged as potential relatives or duplicates and removed as appropriate. If we identify related subjects, we will endeavor to maintain cases. If cases are related, we will randomly select the case to exclude.

Population stratification: Because the CCSS discovery GWAS will be restricted to individuals of European
The impact of selected SNPs will be investigated with high accuracy. Potential pitfalls and alternative strategies: We have developed a staged approach that promises a high likelihood of success. For Aims 1.1 and 1.2, the analytic approach will be similar, but the number of variants being evaluated will differ. After QC, we will detect marginal and GxE interaction effect for all retained SNPs using Cox regression techniques. Therapeutic exposures and cardiovascular risk factors will be adjusted as confounders, based on potential association of these variables with CD/CHF. We will use linkage disequilibrium (LD)-based SNP pruning to estimate the number of independent tests, and use this to correct for multiple testing. To test GxE effect, we will employ Cox regression model as incorporating GxE as a product term and adjusting for age at primary cancer, sex, anthracycline exposure and SNP genotypes. Genotype coding will be additive, dominant and recessive. We will also explore analyses stratified by cumulative anthracycline exposure (<250 mg/m²; ≥250mg/m²) Power (GWAS): Given α as 5x10⁻⁸, and 220 cases/5096 controls, we have >95% power to detect the main effect of a risk variant in additive inheritance with MAF of 30% and hazard ratio (HR)=2.4; we have >85% power to detect GxE effect of a risk variant in dominant inheritance with MAF of 20% and OR=10. These GxE effect sizes are supported by previous studies for CD/CHF (OR=10.9(23), OR=39.1(17), OR=10.2(16)). Power (Candidate gene approach): With customized SNP arrays (for 220 cases), we anticipate that we will have >95% power to detect SNPs with main effect (MDOR=2.1 and MAF >0.2; α=5x10⁻⁴). We will have >90% power to identify GxE interaction >4 (MAF >0.2; α=5x10⁻⁴). Replication (ALTE03N1): Using conditional logistic regression, we will replicate the top SNPs (p< 5 x10⁻⁶) in Aim 1.1, adjusting for age at cancer diagnosis, sex, anthracycline dose, cardiovascular risk factors and radiation to the chest. The cases and controls are matched on primary cancer diagnosis, year of diagnosis and race/ethnicity; the controls are followed for a period that is longer than the latency for development of the cardiac event in the index case. All cases (and controls) that overlap with the CCSS cohort will be excluded from the analysis.

Analytic approach: Aim 1: Identify genetic variants associated with anthracycline-associated CD/CHF in childhood cancer survivors: For Aims 1.1 and 1.2, the analytic approach will be similar, but the number of variants being evaluated will differ. After QC, we will detect marginal and GxE interaction effect for all retained SNPs using Cox regression techniques. Therapeutic exposures and cardiovascular risk factors will be adjusted as confounders, based on potential association of these variables with CD/CHF. We will use linkage disequilibrium (LD)-based SNP pruning to estimate the number of independent tests, and use this to correct for multiple testing. To test GxE effect, we will employ Cox regression model as incorporating GxE as a product term and adjusting for age at primary cancer, sex, anthracycline exposure and SNP genotypes. Genotype coding will be additive, dominant and recessive. We will also explore analyses stratified by cumulative anthracycline exposure (<250 mg/m²; ≥250mg/m²)  

Power (GWAS): Given α as 5x10⁻⁸, and 220 cases/5096 controls, we have >95% power to detect the main effect of a risk variant in additive inheritance with MAF of 30% and hazard ratio (HR)=2.4; we have >85% power to detect GxE effect of a risk variant in dominant inheritance with MAF of 20% and OR=10. These GxE effect sizes are supported by previous studies for CD/CHF (OR=10.9(23), OR=39.1(17), OR=10.2(16)).  

Power (Candidate gene approach): With customized SNP arrays (for 220 cases), we anticipate that we will have >95% power to detect SNPs with main effect (MDOR=2.1 and MAF >0.2; α=5x10⁻⁴). We will have >90% power to identify GxE interaction >4 (MAF >0.2; α=5x10⁻⁴). Replication (ALTE03N1): Using conditional logistic regression, we will replicate the top SNPs (p< 5 x10⁻⁶) in Aim 1.1, adjusting for age at cancer diagnosis, sex, anthracycline dose, cardiovascular risk factors and radiation to the chest. The cases and controls are matched on primary cancer diagnosis, year of diagnosis and race/ethnicity; the controls are followed for a period that is longer than the latency for development of the cardiac event in the index case. All cases (and controls) that overlap with the CCSS cohort will be excluded from the analysis.

Analytic approach: Aim 2: Define functional relevance of genetic variants. We will generate human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CM models with knockouts of the key genetic variants (e.g., CBR3, HAS3, CELF4, others) using CRISPR-Cas9 and will compare anthracycline-related CD/CHF to isogenic cells. For genes that validate, we will perform SNP correction in hiPSCs from patients both with variant of interest and CD/CHF to confirm that the SNP previously identified is the effector. Over 130 samples of human myocardium (open procurement) will be used to a) document the presence of genotype-phenotype associations in cardiac tissue, and b) examine the extent of inter-individual variably in candidate gene expression at the mRNA (q-rtPCR) and protein levels (nano-LC/triple-quad MS). The functional and structural impact of selected SNPs will be investigated with in-vitro assays using recombinant proteins (enzyme kinetics, differential titration calorimetry), substrates, and molecular modeling.

Potential pitfalls and alternative strategies: We have developed a staged approach that promises a high likelihood of success. For Aim 1, we will begin with an agnostic approach, with availability of a well-annotated sample for replication. In addition, in the event that no genetic variant is identified as part of this approach, we are selecting candidate SNPs that have been identified and replicated in previously-published CD/CHF studies. Our detectable effect size is in keeping with previous assessments. For Aim 2, we aim to develop a risk prediction model that will perform better that the clinical model. Again, we have plans for replication of the model. For Aim 3, we have developed several avenues to determine the functional relevance of the identified genetic variants, using both iPSC and CRISPR mechanism as well as myocardial tissue.

Timeline and funding: Table 3 presents the timeline for the study. We have applied for funding through the PA-17-239 (R01) mechanism to complete this project (February, 2018). If external funding is not obtained, we will use internal resources and existing resources to conduct targeted replication and validation.
References: