VOLUME 34 · NUMBER 8 · MARCH 10, 2016

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

CELF4 Variant and Anthracycline-Related Cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study

Xuexia Wang, Can-Lan Sun, Adolfo Quiñones-Lombraña, Purnima Singh, Wendy Landier, Lindsey Hageman, Molly Mather, Jerome I. Rotter, Kent D. Taylor, Yii-Der Ida Chen, Saro H. Armenian, Naomi Winick, Jill P. Ginsberg, Joseph P. Neglia, Kevin C. Oeffinger, Sharon M. Castellino, Zoann E. Dreyer, Melissa M. Hudson, Leslie L. Robison, Javier G. Blanco, and Smita Bhatia

Author affiliations appear at the end of this article.

Published online ahead of print at www.jco.org on January 25, 2016.

Supported by the Children's Oncology Group National Clinical Trials Network Operations Grant No. U10CA180886 (to P. Adamson), The Leukemia and Lymphoma Society Grant No. 6093-08 (S.B.), Grant No. GM073646 (J.G.B.) from State University of New York at Buffalo, National Institutes of Health, National Institute for General Medical Sciences Pharmacogenomics Research Network Grant No. U01 GM92666 (M.Relling), American Lebanese Syrian Associated Charities, Grant No. P30CA033572 (M.Relling), and Grant No. UL1 R0000124 (M.Relling) from St Jude Children's Research Hospital.

X.W., C.-L. S., and A. Q.-L. contributed equally to this work.

Presented in part at the 51st American Society of Clinical Oncology Annual Meeting, Chicago, IL, May 30-June 2, 2015, and at the 14th International Conference on Long-Term Complications of Treatment of Children and Adolescents for Cancer, Alexandria, VA, June 11-13, 2015.

Authors' disclosures of potential conflicts of interest are found in the article online at www.jco.org. Author contributions are found at the end of this article.

Corresponding author: Smita Bhatia, MD, MPH, Institute for Cancer Outcomes and Survivorship, School of Medicine, University of Alabama at Birmingham, 1600 7th Ave S, Lowder 500, Birmingham, AL 35233; e-mail: sbhatia@peds.uab.edu.

0732-183X/16/3408w-863w/$20.00

DOI: 10.1200/JCO.2015.63.4550

ABSTRACT

Purpose
Interindividual variability in the dose-dependent association between anthracyclines and cardiomyopathy suggests that genetic susceptibility could play a role. The current study uses an agnostic approach to identify genetic variants that could modify cardiomyopathy risk.

Methods
A genome-wide association study was conducted in childhood cancer survivors with and without cardiomyopathy (cases and controls, respectively). Single-nucleotide polymorphisms (SNPs) that surpassed a prespecified threshold for statistical significance were independently replicated. The possible mechanistic significance of validated SNP(s) was sought by using healthy heart samples.

Results
No SNP was marginally associated with cardiomyopathy. However, SNP rs1786814 on the *CELF4* gene passed the significance cutoff for gene-environment interaction ($P_{ge} = 1.14 \times 10^{-5}$). Multivariable analyses adjusted for age at cancer diagnosis, sex, anthracycline dose, and chest radiation revealed that, among patients with the A allele, cardiomyopathy was infrequent and not dose related. However, among those exposed to greater than 300 mg/m^2 of anthracyclines, the rs1786814 GG genotype conferred a 10.2-fold (95% CI, 3.8- to 27.3-fold; $P < .001$) increased risk of cardiomyopathy compared with those who had GA/AA genotypes and anthracycline exposure of 300 mg/m^2 or less. This gene-environment interaction was successfully replicated in an independent set of anthracycline-related cardiomyopathy. CUG-BP and ETR-3-like factor proteins control developmentally regulated splicing of *TNNT2*, the gene that encodes for cardiac troponin T (cTnT), a biomarker of myocardial injury. Coexistence of more than one cTnT variant results in a temporally split myofilament response to calcium, which causes decreased contractility. Analysis of *TNNT2* splicing variants in healthy human hearts suggested an association between the rs1786814 GG genotype and coexistence of more than one *TNNT2* splicing variant (90.5% GG *v* 41.7% GA/AA; $P = .005$).

Conclusion
We report a modifying effect of a polymorphism of *CELF4* (rs1786814) on the dose-dependent association between anthracyclines and cardiomyopathy, which possibly occurs through a pathway that involves the expression of abnormally spliced *TNNT2* variants.

J Clin Oncol 34:863-870. © 2016 by American Society of Clinical Oncology

INTRODUCTION

Anthracyclines are a highly effective class of chemotherapeutic agents that play a critical role in the first-line treatment of childhood cancer. However, the strong dose-dependent association with cardiomyopathy[1,2] limits therapeutic potential. Interindividual variability in the dose-dependent association between anthracycline exposure and cardiomyopathy risk[3] suggests that genetic variants possibly modify this association. Previous efforts led by our team and other investigators have used a candidate gene approach to understand the molecular pathogenesis of anthracycline-related cardiotoxicity, and these efforts have resulted in biologically plausible leads.[1,4-8] However, there is a paucity of studies that attempt a comprehensive

agnostic approach with independent replication or functional validation in a robust sample of childhood cancer survivors and with a specific emphasis on the gene-environment (anthracycline dose) interaction. We addressed this gap by performing a genome-wide association study (GWAS) to identify novel single-nucleotide polymorphisms (SNPs) that potentially modified the dose-dependent risk of anthracycline-related cardiomyopathy. SNPs that surpassed a prespecified threshold for statistical significance in the discovery stage were validated in the replication stage by using an independent set of patients who had cardiomyopathy. The possible mechanistic significance of validated SNP(s) was evaluated with healthy heart samples.

METHODS

Study Design and Population

Discovery set. Study participants were drawn from a Children's Oncology Group study (COG-ALTE03N1; NCT00082745) that used a matched case-control design to understand the pathogenesis of cardiomyopathy in childhood cancer survivors. COG member institutions (Data Supplement) enrolled patients after approval was obtained from local institutional review boards. Written informed consent or assent was obtained from patients, parents, or legal guardians. Cases and controls were identified from individuals diagnosed with cancer at age 21 years or younger. Cases consisted of childhood cancer patients who developed cardiomyopathy and who were alive at study participation. For each case, one to four patients who had no signs or symptoms of cardiomyopathy were randomly selected as controls from the same COG childhood cancer survivor cohort and were matched on primary cancer diagnosis, year of diagnosis (within 5 years), and race/ethnicity. The selected controls also needed to have a longer duration of cardiomyopathy-free follow-up compared with time from cancer diagnosis to cardiomyopathy for the corresponding case. The discovery set included 430 childhood cancer survivors (162 cases, 268 controls). All participants provided a biologic specimen (blood [89%] or buccal cells/saliva [11%]) for DNA.

Replication set. An independent set of 54 childhood cancer survivors with cardiomyopathy (enrolled in COG-ALTE03N1 after completion of the discovery stage) comprised the replication set.

Phenotype Assessment

Anthracycline-exposed participants had normal cardiac function before anthracycline exposure. Cases fulfilled the American Heart Association criteria for cardiac compromise[9] by presentation with signs and/or symptoms (dyspnea, orthopnea, fatigue, edema, hepatomegaly, and/or rales) or, in the absence of signs and/or symptoms, with echocardiographic features of left ventricular dysfunction (ejection fraction [EF] 40% or less and/or fractional shortening [SF] 28% or less; Data Supplement).

Therapeutic Exposures

Lifetime anthracycline exposure was calculated by multiplying the cumulative dose (mg/m^2) of individual anthracyclines received by a factor that reflects the cardiotoxic potential of the drug[10] and then by summing the results. Radiation to the chest with heart in the field was captured as a yes/no variable.

Genotyping and Quality Control

Discovery set. Genomic DNA was isolated from peripheral blood (QIAamp kits; Qiagen, Hilden, Germany), buccal cells (Puregene kits; Qiagen), or saliva (Oragene kits, Ottowa, Ontario, Canada). Genotyping was performed on the Illumina HumanOmniExpress-12 v1.0 DNA analysis bead-chip (Illumina, San Diego, CA) at the Medical Genetics Institute of Cedars-Sinai. Quality control (QC) tests for genotype data were performed with PLINK.[11] We did not identify any individuals on the basis of discordant sex, sample contamination, or low genotyping (missing fraction $>$ 0.025). We filtered 60 individuals on the basis of results of the heterozygosity test (inbreeding coefficient $|$ F $|$ was $>$ 0.1[12]) and two individuals on the basis of the results of the relatedness test (identity by descent value $>$ 0.1875). The multidimensional scaling method[11] was used to cluster individuals in the discovery set into non-Hispanic whites and other. To control for potential population stratification, 37 individuals were filtered, which retained 331 non-Hispanic whites (112 patients as cases, 219 patients as controls) in the discovery stage.

Both the overall genomic control inflation factor (λ = 1.04) and the quantile-quantile plot (Data Supplement) for genome-wide marginal effect tests did not suggest any large-scale systematic bias as a result of population stratification. The genotype data included 709,358 autosomal SNPs. We removed 2,999 SNPs that failed the missing data test (missing fraction $>$ 0.05) and 119,316 SNPs that had a minor allele frequency (MAF) of less than 0.05. We also checked the Hardy-Weinberg equilibrium and excluded 3,295 SNPs that had *P* values less than .0001. The final data set included 583,748 autosomal SNPs (82.3%) for each of the 331 individuals and had a total genotyping call rate of 99.8%.

Replication set. Genomic DNA was isolated from peripheral blood (QIAamp kits; Qiagen), buccal cells (Puregene kits) or saliva (Oragene kits). Significant SNP(s) identified in discovery were genotyped in the replication set by using Sequenom iPLEX SNP chemistry on a MassArray system.

Functional Analysis

Healthy heart samples (n = 33) were procured from the National Disease Research Interchange and the Cooperative Human Tissue Network. The State University of New York at Buffalo's Institutional Review Board approved this research. Presence of SNP(s) of interest in cardiac DNA was investigated with TaqMan genotyping assays per the manufacturer's guidelines (Applied Biosystems, Carlsbad, CA). Splicing isoforms were amplified by nested polymerase chain reaction (PCR). Identities of PCR products were verified by DNA sequencing (Data Supplement).

Statistical Analyses

Analyses were conducted using R (http://www.r-project.org/), SAS software (SAS Institute, Cary, NC), and PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/).

Discovery stage.

Marginal effect. Conditional logistic regression techniques (model 1) were used for each SNP (n = 583,748) that passed QC, MAF, and Hardy-Weinberg equilibrium filters. Age at diagnosis of primary cancer, sex, chest radiation, and anthracycline dose were included in the analysis on the basis of previous evidence about the association of these variables with cardiomyopathy.[13]

$$\text{logit}\,(p) = \text{matched\_set} + \text{agedx} + \text{gender} + \text{RT} + \text{anthracycline} + \text{SNP} \quad \text{(model 1)}$$

In model 1, p is the probability of cardiomyopathy conditional on matched set; SNP is the genotype for each SNP in additive coding; anthracycline is the cumulative anthracycline dose in mg/m^2 (continuous variable); RT is chest radiation (yes/no); agedx is the age at diagnosis of primary cancer (continuous variable); gender is sex of participants (male/female).

Gene-environment interaction. We used a two-step procedure[14-17] to maximize our ability to detect a role for gene-environment (anthracycline dose) interaction in the development of anthracycline-related cardiomyopathy. Step 1 involved retention of suggestive SNPs that had modest main effects. Step 2 involved testing the SNPs retained in step 1 for their roles in modification of the association between anthracycline dose and cardiomyopathy risk (ie, gene-environment interactions). Independence between the tests in the two steps has been proven.[18]

In Step 1, conditional logistic regression techniques (using model 1) were used for each SNP, and 1,000 suggestive SNPs (corresponding to $P < .004$)

were retained for step 2 (Data Supplement). In step 2, gene-environment interaction analysis was conducted with conditional logistic regression techniques (model 2) for all retained SNPs. The variable legend for model 2 is the same as that of model 1.

logit (p) = matched_set + agedx + gender + RT + anthracycline + SNP + SNP* anthracycline
(model 2)

After taking into consideration linkage disequilibrium (LD) among the 1,000 retained SNPs, 643 independent tests (variance inflation factor = 2) were estimated by the repeated sliding-window procedure.[11] This allowed a P value less than 7.77×10^{-5} (0.05/643) to serve as the threshold for the whole-genome significance test after accounting for multiple testing.[12]

Replication stage. We used a case-only design to verify significant gene-environment interactions identified in the discovery stage. Cumulative anthracycline exposure was dichotomized as low-to-moderate-dose ($\leq$ 300 mg/m^2) and high-dose ($>$ 300 mg/m^2) on the basis if previous observations[1,19] of elevated cardiomyopathy risk with a dose greater than 300 mg/m^2. We treated the binary variable of anthracycline exposure as a dependent variable and used logistic regression techniques to conduct a gene-environment interaction analysis (model 3).

logit (p_anth_exp) = agedx + gender + RT + race/ethnicity + SNP
(model 3)

In model 3, p_anth_exp is the probability of being in the high-dose anthracycline group; anth_exp is dichotomized anthracycline exposure (0: $\leq$ 300 mg/m^2; 1: $>$ 300 mg/m^2). The rest of the variable legend for model 3 is the same as that of model 1.

Table 1. Characteristics of Study Population in the Discovery and Replication Set

	Discovery Set			Replication Set	
	No. (%) of Non-Hispanic Whites			No. (%) of Cases	
Variable	Cases (n = 112)	Controls (n = 219)	*P**	Non-Hispanic Whites (n = 21)	All Races/ Ethnicities (n = 54)
Race/ethnicity†					
Non-Hispanic white	112 (100)	219 (100)	Matched	21 (100)	21 (38.9)
Hispanic	0 (0)	0 (0)		0 (0)	12 (22.2)
Other	0 (0)	0 (0)		0 (0)	21 (38.9)
Age at primary cancer diagnosis, years					
Mean (SD)	8.4 (5.7)	8.3 (5.8)	.75	7.5 (4.8)	7.7 (5.0)
Median (range)	7.5 (0-20)	7.9 (0-21)		5.6 (1.05-17.27)	7.7 (0.02-20.6)
Age at study participation, years					
Mean (SD)	19.0 (9.5)	21.9 (9.1)	< .001	19.3 (7.8)	18.0 (8.5)
Median (range)	18.3 (0.4-41.8)	20.5 (4.2-50.1)		17.7 (3.9-35.6)	16.6 (1.1-40.4)
Female sex	66 (58.9)	113 (51.6)	.24	12 (57.1)	24 (44.4)
Primary diagnosis†					
HL/NHL	28 (25.0)	49 (22.4)	Matched	2 (9.6)	10 (18.5)
Sarcoma	35 (31.2)	61 (27.9)		5 (23.8)	17 (31.5)
ALL/AML	27 (24)	67 (30.6)		10 (47.6)	18 (33.2)
Other	22 (19.6)	42 (19.1)		4 (19.0)	9 (16.7)
Year of primary cancer diagnosis†					
≤ 1990	46 (41)	71 (32.4)	Matched	7 (33.3)	18 (33.3)
1991-2000	47 (42)	101 (46.1)		8 (38.1)	21 (38.9)
2001-2008	19 (17)	47 (21.5)		6 (28.6)	15 (27.8)
Median (range) length of follow-up, years	9.4 (0.1-35.1)	12.9 (1.4-41)	< .001	8.1 (0.8-30.6)	4.1 (0.1-30.6)
Cumulative anthracycline exposure, mg/m^2‡					
Median (range)	319 (0-760)	180 (0-825)	< .001	350 (0-668)	301 (0-668)
> 300	59 (52.7)	76 (34.7)		14 (67)	27 (50.3)
Received chest radiation§	25 (22.3)	27 (12.4)	.03	11 (52)	19 (35)
Median (range) age at cardiomyopathy diagnosis, years	18.3 (0.4-41.7)	NA		15.3 (3.9-34.5)	14.5 (1.1-34.6)
Median ejection fraction, %	44 (13-68)‖	65 (47-84)¶	—	37 (32-55)#	41 (20-60)#
Median fractional shortening, %	24 (5-33)[a]	37 (28-47)[b]		18.5 (14-41)	21 (10-41)
CELF4 genotype					
GG	78 (69.6)	123 (56.2)	.02	17 (81)	41 (75.9)
GA	32 (28.6)	83 (37.9)		4 (19)	13 (24.1)
AA	2 (1.8)	13 (5.9)		0 (0)	0 (0)

Abbreviations: ALL, acute lymphocytic leukemia; AML, acute myelogenous leukemia; HL, Hodgkin lymphoma; NA, not applicable; NHL, non-Hodgkin lymphoma; SD, standard deviation.
*P values were obtained from conditional logistic regression or generalized linear model, taking into consideration matched set.
†Matching variables. Because of variation in the number of controls per case, the percentage of controls and cases in each category of a specific matching variable may not be identical.
‡Missing anthracycline dose for one control.
§n = 15 patients with no exposure to anthracyclines received radiation (n = 5 cases, n = 10 controls).
‖n = 89.
¶n = 79.
#n = 9.
[a]n = 93.
[b]n = 111.

RESULTS

Discovery Stage

Demographic/Clinical Characteristics. The median ages at primary cancer diagnosis for patients in the cases and controls were 7.5 and 7.9 years, respectively (Table 1). Cases received a higher cumulative anthracycline exposure (median dose, 319 mg/m^2 *v* 180 mg/m^2; $P < .001$) and were more likely to have received chest radiation (22.3% *v* 12.4%; $P = .03$). The median time for the case group between cancer diagnosis and cardiomyopathy was 9.4 years; controls were followed for a significantly longer period (median, 12.9 years; $P < .001$).

Cases met the criteria for cardiomyopathy (median EF, 44% [range, 13% to 68%]; median SF, 24% [range, 5% to 33%]; Data Supplement). Controls had no symptoms or signs of cardiac compromise at study participation. Of the 219 controls, 137 had normal echocardiographic features (median EF, 65%; median SF, 37%); the remaining 82 controls did not have echocardiograms. Thirty-one of the 82 controls without echocardiograms did not receive cardiotoxic drugs, whereas the remaining 51 received cardiotoxic agents (anthracyclines alone, n = 38; radiation alone, n = 7; both, n = 6). Exclusion of the 51 anthracycline- and/or radiation-exposed controls without echocardiograms from the analysis did not alter the results (Data Supplement); we opted to include them in the analysis.

Risk of cardiomyopathy. Multivariable conditional logistic regression analysis, adjusted for age at diagnosis of primary cancer and sex, revealed that patients exposed to high-dose anthracyclines (> 300 mg/m^2) were 5.1 times (95% CI, 2.4 to 10.9 times; $P < .001$) more likely to develop cardiomyopathy than those exposed to 300 mg/m^2 or less. Exposure to chest radiation was associated with a 3.2-fold increased risk of cardiomyopathy (95% CI, 1.3- to 7.7-fold; $P = .01$) compared with no radiation. No SNP was marginally associated with cardiomyopathy at the genome-wide level.

Gene-environment interaction analysis. One SNP (rs1786814; MAF in cases = 0.18, MAF in controls = 0.25) in the gene CUGBP Elav-like family member 4 (*CELF4*) on chromosome 18 exceeded the multiple-comparison–corrected threshold (7.77×10^{-5}) for a significant SNP-by-anthracycline interaction in the two-step method ($P = 1.14 \times 10^{-5}$; Fig 1).

The association between SNP rs1786814 and cardiomyopathy risk is shown in Table 2 and Table 3. After adjustment for age at primary cancer diagnosis, sex, chest radiation, and anthracycline dose, the GG genotype was associated with an increased risk of cardiomyopathy compared with the GA/AA genotype (odds ratio [OR], 2.26; 95% CI, 1.2 to 4.0; $P = .006$). Furthermore, patients who had the GG genotype who received high-dose anthracyclines (> 300 mg/m^2) were at a 10.16-fold increased risk of cardiomyopathy (95% CI, 3.8- to 27.3-fold; $P < .001$) compared with those who had the GA/AA genotype and low-to-moderate dose exposure (≤ 300 mg/m^2). The modifying effect of the rs1786814 genotype on the dose-dependent association between anthracycline and cardiomyopathy risk is also shown in Figure 2. Among individuals with GA or AA genotypes, cardiomyopathy risk did not increase with increased anthracycline exposure. However, among individuals with GG genotype, cardiomyopathy risk increased dramatically for those exposed to greater than 300 mg/m^2 of anthracyclines.

Replication stage

Demographic and clinical characteristics. The median age at primary cancer diagnosis of the 54 patients who had cardiomyopathy was 7.7 years; 39% of the patients were non-Hispanic whites (Table 1). The median cumulative anthracycline exposure was 301 mg/m^2; the median EF was 41%, and the median SF was 21%.

Gene-environment interaction. Compared with cases who had GA/AA genotype, the odds of cases who had the GG genotype being in the greater-than–300-mg/m^2 versus the 300-mg/m^2 –or-less anthracycline group were 5.09 times higher (95% CI, 1.0 to 25.2 times; $P = .046$) for the entire replication set (Table 2 and Table 3). In addition, the odds of cases who had the GG genotype being in the

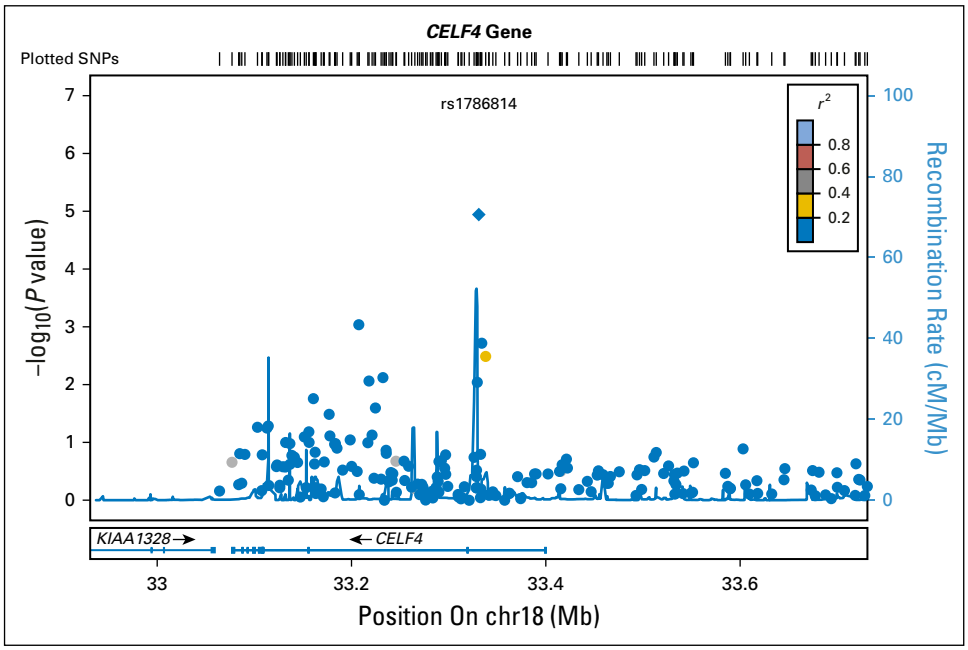

Fig 1. P values of gene-by-environment interactions for the gene *CELF4* region. Gene *CELF4* and adjacent genes in the region are shown in the lower panel, and the unbroken blue line indicates the recombination rate within the region. Each filled circle represents the *P* value of a gene-by-environment interaction for one single-nucleotide polymorphism (SNP) in the non-Hispanic whites discovery set. SNPs in the region were colored according to their degrees of correlation (r^2) with rs1786814, which was estimated internally by LocusZoom[38] on the basis of Utah residents of Northern and Western European ancestry (EUR) haplotypes of 1,000 genomes.

Table 2. Main and Modifying Effect of *CELF4* rs1786814 Genotypes on Dose-Dependent Risk of Anthracycline-Related Cardiomyopathy in the Discovery Set

CELF4 rs1786814 genotype status	No. (%) in Discovery Set: Cases (n = 112)	Controls (n = 219)	Odds Ratio (95% CI)	*P*
Main effect*				
AA	2 (2)	13 (6)	1.0	—
GA	32 (28)	83 (38)	3.18 (0.5 to 18.9)	.2
GG	78 (70)	123 (56)	6.73 (1.1 to 41.1)	.04
AA and GA	34 (30)	96 (44)	1.00	—
GG	78 (70)	123 (56)	2.26 (1.2 to 4.0)	.006
CELF4 gene–environment (anthracycline) interaction effect†				
Cumulative anthracycline exposure and genotype status				
≤ 300 mg/m^2 and GA/AA	19 (17)	52 (24)	1.0	—
≤ 300 mg/m^2 and GG	34 (30)	90 (41)	1.03 (0.5 to 2.2)	.9
> 300 mg/m^2 and GA/AA	15 (14)	43 (20)	1.75 (0.6 to 5.0)	.3
> 300 mg/m^2 and GG	44 (39)	33 (15)	10.16 (3.8 to 27.3)	< .001

*Odds ratios were obtained from conditional logistic regression that adjusted for age at diagnosis, sex, chest radiation, anthracycline exposure (continuous).
†Odds ratios were obtained from conditional logistic regression that adjusted for age at diagnosis, sex, and chest radiation.

high-dose anthracycline category were also elevated for non-Hispanic whites (OR, 14.59; *P* = .07) and for non-Hispanic whites and Hispanics together (OR, 13.74; *P* =.04).

Imputation Analysis

SNP rs1786814 is located in the intronic region of *CELF4* on chromosome 18 (chr18: 35077028), and it appears to be in a region of low LD; LD between rs1786814 and other SNPs on *CELF4* was not strong (Fig 1). We imputed the entire chromosome 18 using 1,000 genome SNPs in 566 EUR (Utah residents of Northern and Western European ancestry) –phased haplotypes as a reference (Data Supplement).[20] No significant gene-environment interaction was identified in the analysis on the basis of imputed SNPs and r^2 greater than 0.3.

rs1786814 Genotype Status and Cardiac Expression of TNNT2 *Splicing Variants*

The CELF family of RNA-binding proteins is implicated in the alternative splicing of the *TNNT2* gene during development.[21,22] *TNNT2* encodes for troponin T, which is located on the thin filament of striated muscles and which regulates muscle contraction in response to alterations in intracellular calcium ion (Ca^{2+}) concentration.[19] Coexistence of the embryonic and adult *TNNT2* variants results in a temporally split myofilament response to increasing intracellular Ca^{2+} concentrations, which causes decreased ventricular pumping efficiency.[20,21] We investigated whether the *CELF4* rs1786814 genotype status is associated with expression of *TNNT2* variants in myocardial tissue. The following genotype distribution for the *CELF4* SNP rs1786814 was found in 33 samples of cardiac DNA: GG (63.6%), GA (24.2%), and AA (12.2%). The embryonic splicing variant of *TNNT2* was found in 24 of the 33 heart samples and was more likely to coexist with the adult *TNNT2* splicing variant in hearts homozygous for the high-risk rs1786814 genotype (GG: 19 (90.5%) of 21 samples; GA/AA: 5 (41.7%) of 12 samples; *P* = .005, Fisher's exact test; Data Supplement).

Replication of Previous Findings

We have previously reported the dose-modifying impact of *CBR3*[1] and *HAS3*[7] on anthracycline-related cardiomyopathy risk. In the current study (84% overlap with a previous study[1]), we demonstrate that the *CBR3*:GG genotype is associated with an 8.8-fold increased cardiomyopathy risk compared with *CBR3*:GA/AA genotypes (*P* = .03) in individuals exposed to low-to-moderate doses of anthracyclines. By using the ITMAT/Broad CARe (IBC; Illumina) cardiovascular SNP array that profiles SNPs in 2,100 genes considered relevant for cardiovascular disease in the general population,[23] we reported that, among those exposed to high-dose anthracyclines, the AA genotype for SNP rs2232228 in the *HAS3* gene conferred an 8.9-fold increased cardiomyopathy risk compared with the GG genotype.[7] We examined the 13,219 overlapped SNPs between the IBC array and the GWAS array for all 331 patients in the discovery set of the current study population. *HAS3* rs2232228 was not retained in the top 1,000 SNPs (*P* = .6). However, the gene-by-environment effect was significant (*P* < .001). Additionally, among the 199 individuals who

Table 3. Main and Modifying Effect of *CELF4* rs1786814 Genotypes on Dose-Dependent Risk of Anthracycline-Related Cardiomyopathy in the Replication Set

CELF4 rs1786814 genotype status for all patients in the replication set	No. (%) of Patients by Cumulative Anthracycline Exposure (mg/m^2): ≤ 300	> 300	Odds Ratio* (95% CI)	*P*
GA and AA	9 (33)	4 (15)	1.0	—
GG	18 (67)	23 (85)	5.09 (1.03 to 25.23)	.046

*Odds ratios were obtained from logistic regression that adjusted for age at diagnosis, primary cancer diagnosis, year of diagnosis, follow-up time, sex, and chest radiation.

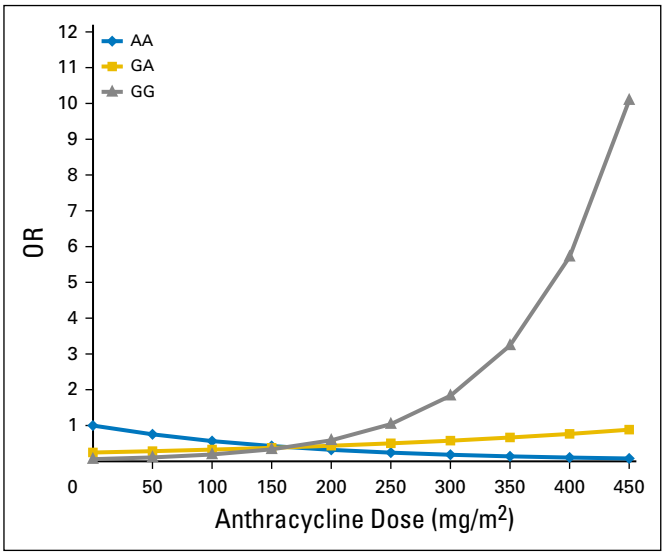

Fig 2. Risk of cardiomyopathy by anthracycline dose and genotype status (AA, GA, GG). Odds ratios (ORs) were calculated based on model 2, which treated anthracycline dose as a continuous variable (reference group: AA genotype with no anthracycline exposure).

had complete genotype data on IBC and GWAS SNPs (n = 605,441), neither *CELF4* SNP rs1786814 (P_{ge} = .006; P_{main} = .01) nor *HAS3* SNP rs2232228 (P_{ge} = .002; P_{main} = .49) passed the significance threshold, likely because of the small sample size.

DISCUSSION

We used a genome-wide approach to study the modifying effect of genetic variants on the dose-dependent association between anthracycline exposure and cardiomyopathy risk. Among individuals with the GA or AA genotype for SNP rs1786814 on the *CELF4* gene, cardiomyopathy risk was not elevated, irrespective of cumulative anthracycline exposure. However, among non-Hispanic white individuals who had the GG genotype, exposure to greater than 300 mg/m^2 of anthracyclines conferred a 10.2-fold increased risk of cardiomyopathy compared with those who had the GA/AA genotype and who were exposed to 300 mg/m^2 or less of anthracyclines. The successful replication of findings in a demographically diverse population speaks to the robustness of the findings in this study. Furthermore, we imputed the entire chromosome 18 and found no additional gene-environment interactions that approached significance. These data implicate the *CELF4* SNP (rs1786814) in the development of anthracycline-related cardiomyopathy.

The CELF protein family belongs to a group of splicing regulators that controls developmentally regulated, tissue-specific splicing events.[21] Analysis of the *CELF4* sequence with Netgene2 software (http://www.cbs.dtu.dk/services/NetGene2/)[24] showed a potential splice donor site for the G allele of rs1786814 (confidence, 55%); the A allele caused the loss of donor splice site. Furthermore, cardiac expression of truncated forms of *CELF4* in mice repress the alternative splicing activity of CELF proteins, which results in extensive cardiac fibrosis and cardiac dysfunction.[21] One of the classic targets of CELF family members is *TNNT2*, the gene encoding for cardiac troponin T (cTnT).[25,26] cTnT has an essential role in Ca^{2+} signaling in cardiac muscle and is an established biochemical marker of myocardial injury[27]; serum cTnT levels are increased after anthracycline treatment.[28] During development, splicing of *TNNT2* leads to the inclusion of alternative exon 5, which results in the insertion of 10 additional amino acids into the protein.[29] mRNA expression of the *TNNT2* splicing variants that carry exon 5 clearly predominates in embryonic hearts but is barely detectable in adult hearts. However, continued expression of embryonic *TNNT2* variant in the adult cardiac muscle could cause coexistence of two or more *TNNT2* variants, which could result in a temporally split myofilament response to increasing Ca^{2+} concentrations and could cause decreased myocardial contractility and ventricular pumping efficiency.[30,31] In fact, coexistence of two or

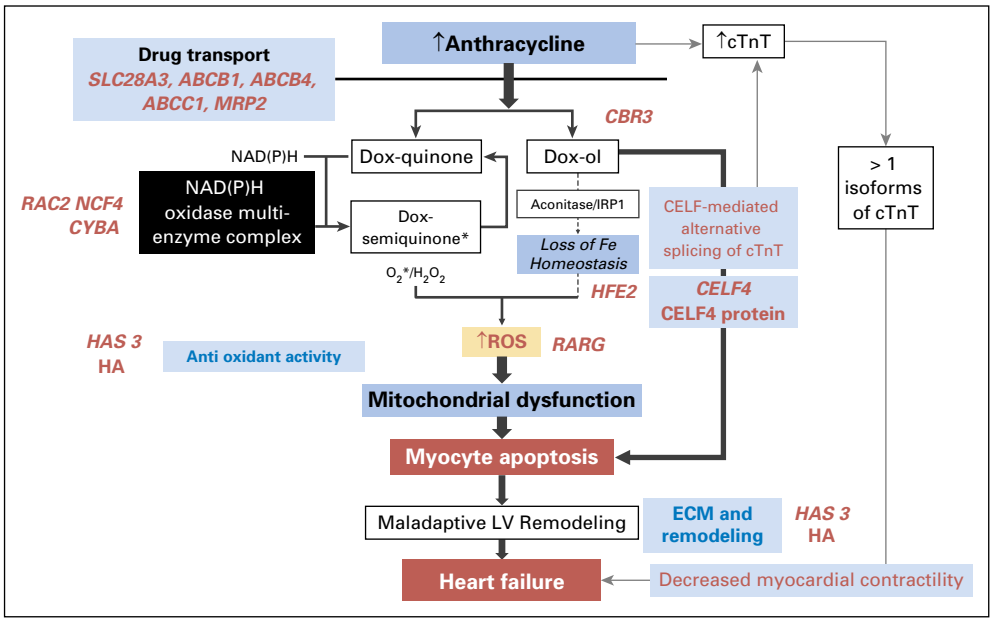

Fig 3. Proposed pathogenesis of anthracycline-related cardiomyopathy. cTnT, cardiac troponin T; Dox-, doxorubicin; HA, hyaluronic acid; IRP1, iron regulatory protein 1; NAD(P)H, nicotinamide adenine dinucleotide phosphate hydrogen; ROS, reactive oxygen species.

more *TNNT2* splicing variants is found in dilated cardiomyopathy.[32] In the current study, we found that heart samples homozygous for the high-risk *CELF4* rs1786814 G allele are more likely to coexpress the embryonic and adult *TNNT2* variants and, thus, to possibly enhance cardiotoxicity risk in patients homozygous for the G allele after exposure to high-dose anthracyclines.

We provide evidence that SNP rs1786814 in the *CELF4* gene alters the risk of anthracycline-related cardiomyopathy among patients exposed to high-dose anthracyclines. For individuals of European descent, the frequency of the SNP rs1786814 G allele is 0.81. The genotype distribution in individuals of European descent is GG (0.66), GA (0.29), and AA (0.04). Thus, with respect to findings from the current study, 66% of childhood cancer survivors are at increased risk for anthracycline-related cardiomyopathy when exposed to high-dose anthracyclines. However, equally as important, 34% of childhood cancer survivors are spared, irrespective of anthracycline dose.

Although this study includes one of the largest populations of clinically validated cardiomyopathy occurrences among childhood cancer survivors, the relatively modest sample size (compared with typical GWASs that focuses on chronic diseases) could have precluded detection of additional genetic associations because of insufficient power. A sample size of 331 patients, as in the current study (with 30% in the case group), achieves 80% power at a .004 significance level for minimally detectable odds ratios that range from 2.5 to 4.3 for MAFs that range from 5% to 50% (Data Supplement).[33] Furthermore, failure of previously reported genetic variants (*HAS3* rs2232228 and *CBR3* V244M) to exceed the multiple-comparison–corrected threshold (7.77×10^{-5}) for a significant SNP-by-anthracycline interaction in the two-step method used in the current study was probably a reflection of the sample size. However, when we examined these candidate SNPs individually, we were able to replicate the findings in the current study. In a typical GWAS, gene-by-environment interactions are investigated by testing the interaction between each SNP and exposure individually, and the findings are adjusted for multiple comparisons.[17,34,35] Although the two-step method enhances statistical power,[14] it does so at the expense of losing some of the susceptibility SNPs in the first step (such as *HAS3* rs2232228 and *CBR3* V244M). This suggests that both the candidate gene approach (which queries a biologically plausible association) and the agnostic genome-wide association approach may have merit. This limitation notwithstanding, findings from the current study as well as previous studies can be integrated to form a unifying model for understanding the pathogenesis of anthracycline-related cardiomyopathy (Fig 3) and for developing a prediction model that would allow identification of patients at highest risk of anthracycline-related cardiomyopathy, so targeted interventions can be instituted.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

AUTHOR CONTRIBUTIONS

Conception and design: Jerome I. Rotter, Leslie L. Robison, Smita Bhatia
Collection and assembly of data:, Prunima Singh, Wendy Landier, Lindsey Hageman, Molly Mather, Jerome I. Rotter, Kent D. Taylor, Yii-Der Ida Chen, Naomi Winick, Jill P. Ginseberg, Joseph P. Neglia, Kevin C. Oeffinger, Sharon M. Castellino, Zoann E. Dreyer, Melissa M. Hudson, Smita Bhatia
Data analysis and interpretation: Xuexia Wang, Can-Lan Sun, Adolfo Quiñones-Lombraña, Purnima Singh, Jerome I. Rotter, Kent D. Taylor, Yii-Der Ida Chen, Saro H. Armenian, Javier G. Blanco, Smita Bhatia
Administrative support:, Wendy Landier, Linsdey Hageman, Molly Mather, Jerome I. Rotter, Smita Bhatia
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Blanco JG, Sun C-L, Landier W, et al: Anthracycline-related cardiomyopathy after childhood cancer: Role of polymorphisms in carbonyl reductase genes—A report from the Children's Oncology Group. J Clin Oncol 30:1415-1421, 2012
2. Grenier MA, Lipshultz SE: Epidemiology of anthracycline cardiotoxicity in children and adults. Semin Oncol 25:72-85, 1998 (suppl 10)
3. Bryant J, Picot J, Levitt G, et al: Cardioprotection against the toxic effects of anthracyclines given to children with cancer: A systematic review. Health Technol Assess 11:iii, ix-x, 1-84, 2007
4. Armenian SH, Ding Y, Mills G, et al: Genetic susceptibility to anthracycline-related congestive heart failure in survivors of haematopoietic cell transplantation. Br J Haematol 163:205-213, 2013
5. Visscher H, Ross CJD, Rassekh SR, et al; Canadian Pharmacogenomics Network for Drug Safety Consortium: Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. J Clin Oncol 30:1422-1428, 2012
6. Wojnowski L, Kulle B, Schirmer M, et al: NAD(P)H oxidase and multidrug resistance protein genetic polymorphisms are associated with doxorubicin-induced cardiotoxicity. Circulation 112:3754-3762, 2005
7. Wang X, Liu W, Sun C-L, et al: Hyaluronan synthase 3 variant and anthracycline-related cardiomyopathy: a report from the children's oncology group. J Clin Oncol 32:647-653, 2014
8. Lipshultz SE, Lipsitz SR, Kutok JL, et al: Impact of hemochromatosis gene mutations on cardiac status in doxorubicin-treated survivors of childhood high-risk leukemia. Cancer 119:3555-3562, 2013
9. Hunt SA, Abraham WT, Chin MH, et al: 2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines—Developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation 119:e391-e479, 2009
10. Shankar SM, Marina N, Hudson MM, et al; Cardiovascular Disease Task Force of the Children's Oncology Group: Monitoring for cardiovascular disease in survivors of childhood cancer—Report from the Cardiovascular Disease Task Force of the Children's Oncology Group. Pediatrics 121:e387-e396, 2008
11. Purcell S, Neale B, Todd-Brown K, et al: PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575, 2007
12. Cappola TP, Li M, He J, et al: Common variants in HSPB7 and FRMD4B associated with advanced heart failure. Circ Cardiovasc Genet 3:147-154, 2010
13. Lipshultz SE, Alvarez JA, Scully RE: Anthracycline associated cardiotoxicity in survivors of childhood cancer. Heart 94:525-533, 2008
14. Kooperberg C, Leblanc M: Increasing the power of identifying gene x gene interactions in genome-wide association studies. Genet Epidemiol 32:255-263, 2008
15. Hsu L, Jiao S, Dai JY, et al: Powerful cocktail methods for detecting genome-wide gene-environment interaction. Genet Epidemiol 36:183-194, 2012
16. Gauderman WJ, Zhang P, Morrison JL, et al: Finding novel genes by testing G × E interactions in a genome-wide association study. Genet Epidemiol 37:603-613, 2013
17. Murcray CE, Lewinger JP, Gauderman WJ: Gene-environment interaction in genome-wide association studies. Am J Epidemiol 169:219-226, 2009

18. Dai JY, Kooperberg C, Leblanc M, et al: Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. Biometrika 99:929-944, 2012

19. Mulrooney DA, Yeazel MW, Kawashima T, et al: Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: Retrospective analysis of the Childhood Cancer Survivor Study cohort. BMJ 339:b4606, 2009

20. 1,000 Genomes Project Consortium: A map of human genome variation from population scale sequencing. Nature 467:1061-1073, 2010

21. Ladd AN, Charlet N, Cooper TA: The CELF family of RNA-binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol 21:1285-1296, 2001

22. Ladd AN, Stenberg MG, Swanson MS, et al: Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. Dev Dyn 233:783-793, 2005

23. Keating BJ, Tischfield S, Murray SS, et al: Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. PLoS One 3: e3583, 2008

24. Brunak S, Engelbrecht J, Knudsen S: Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220:49-65, 1991

25. Horwich TB, Patel J, MacLellan WR, et al: Cardiac troponin I is associated with impaired hemodynamics, progressive left ventricular dysfunction, and increased mortality rates in advanced heart failure. Circulation 108:833-838, 2003

26. Kismet E, Varan A, Ayabakan C, et al: Serum troponin T levels and echocardiographic evaluation in children treated with doxorubicin. Pediatr Blood Cancer 42:220-224, 2004

27. Gordon AM, Homsher E, Regnier M: Regulation of contraction in striated muscle. Physiol Rev 80: 853-924, 2000

28. Kilickap S, Barista I, Akgul E, et al: cTnT can be a useful marker for early detection of anthracycline cardiotoxicity. Ann Oncol 16:798-804, 2005

29. Cooper TA, Ordahl CP: A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. J Biol Chem 260:11140-11148, 1985

30. Biesiadecki BJ, Elder BD, Yu ZB, et al: Cardiac troponin T variants produced by aberrant splicing of multiple exons in animals with high instances of dilated cardiomyopathy. J Biol Chem 277:50275-50285, 2002

31. Gomes AV, Venkatraman G, Davis JP, et al: Cardiac troponin T isoforms affect the Ca(2+) sensitivity of force development in the presence of slow skeletal troponin I: Insights into the role of troponin T isoforms in the fetal heart. J Biol Chem 279: 49579-49587, 2004

32. Anderson PA, Greig A, Mark TM, et al: Molecular basis of human cardiac troponin T isoforms expressed in the developing, adult, and failing heart. Circ Res 76:681-686, 1995

33. Hsieh FY, Bloch DA, Larsen MD: A simple method of sample size calculation for linear and logistic regression. Stat Med 17:1623-1634, 1998

34. Murcray CE, Lewinger JP, Conti DV, et al: Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genet Epidemiol 35:201-210, 2011

35. Lin X, Lee S, Christiani DC, et al: Test for interactions between a genetic marker set and environment in generalized linear models. Biostatistics 14:667-681, 2013

36. Pruim RJ, Welch RP, Sanna S, et al: LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics 26: 2336-2337, 2010

Affiliations

Xuexia Wang, University of Wisconsin-Milwaukee, Milwaukee, WI; Can-Lan Sun, Molly Mather, Saro H. Armenian, City of Hope, Duarte; Jerome I. Rotter, Kent D. Taylor, Yii-Der Ida Chen, Los Angeles Biomedical Research Institute at Harbor–University of California, Los Angeles, Torrance, CA; Adolfo Quiñones-Lombraña, Javier G. Blanco, State University of New York at Buffalo, Buffalo; Kevin C. Oeffinger, Memorial Sloan Kettering Cancer Center, New York, NY; Purnima Singh, Wendy Landier, Lindsey Hageman, Smita Bhatia, University of Alabama at Birmingham, Birmingham, AL; Naomi Winick, University of Texas Southwestern Medical Center, Dallas; Zoann E. Dreyer, Texas Children's Cancer Center, Houston, TX; Jill P. Ginsberg, Childrens Hospital of Philadelphia, Philadelphia, PA; Joseph P. Neglia, University of Minnesota, Minneapolis, MN; Sharon M. Castellino, Emory University and Children's Healthcare of Atlanta, Atlanta, GA; and Melissa M. Hudson, Leslie L. Robison, St Jude Children's Research Hospital, Memphis, TN.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

CELF4 Variant and Anthracycline-Related Cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Xuexia Wang
No relationship to disclose

Can-Lan Sun
No relationship to disclose

Adolfo Quiñones-Lombraña
No relationship to disclose

Purnima Singh
No relationship to disclose

Wendy Landier
Research Funding: Merck Sharp & Dohme (Inst)

Lindsey Hageman
No relationship to disclose

Molly Mather
No relationship to disclose

Jerome I. Rotter
No relationship to disclose

Kent D. Taylor
No relationship to disclose

Yii-Der Ida Chen
No relationship to disclose

Saro H. Armenian
No relationship to disclose

Naomi Winick
No relationship to disclose

Jill P. Ginsberg
No relationship to disclose

Joseph P. Neglia
No relationship to disclose

Kevin C. Oeffinger
No relationship to disclose

Sharon M. Castellino
No relationship to disclose

ZoAnn E. Dreyer
No relationship to disclose

Melissa M. Hudson
No relationship to disclose

Leslie L. Robison
No relationship to disclose

Javier G. Blanco
No relationship to disclose

Smita Bhatia
No relationship to disclose

Acknowledgment

We thank the laboratory assistance provided by Mary Relling, and her research staff, particularly Pamela McGill, Natalie Lowery, Sean Freeman, and Nancy Kornegay. We also thank the patients and families for their participation.