Evaluating the role of genetic variants on blood cell count variability in the Jackson Heart Study

J.R. Shaw[1], R.W. Corty[2], L.M. Raffield[2], A.P. Reiner[3], J.G. Wilson[4], L.A. Lange[1], W. Valdar[2], E.M. Lange[1]

1) University of Colorado Anschutz Medical Campus, Aurora, CO; 2) University of North Carolina, Chapel Hill, NC; 3) University of Washington, Seattle, WA; 4) University of Mississippi, Jackson, MS

Introduction

While genome wide association studies (GWAS) have identified thousands of genetic loci associated with differences in mean values of human quantitative traits (QTL), variants associated with variability around the mean remain relatively uninvestigated in human populations. Using the double-generalized linear model, we performed a genome-wide scan for QTL and vQTL associated with white blood cell count (WBCC) in 2,257 unrelated African American participants from the Jackson Heart Study (JHS).

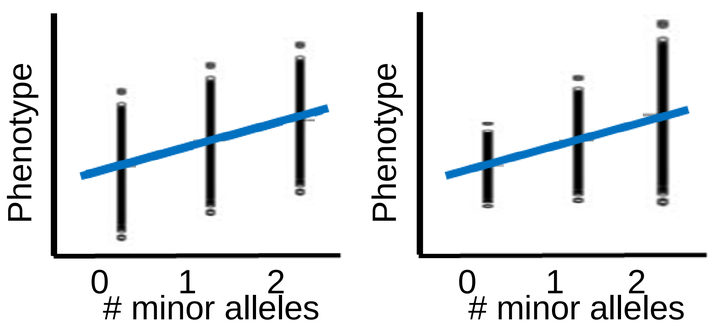
Background

Variability Quantitative Trait Loci (vQTL)

Quantitative trait loci associated with trait variability (vQTL) are variants for which the variability around the mean trait value differs by genotype (Figure 1). Variants associated with trait variability may represent direct genetic influence on phenotypic variability, or indirect influence due to genetic interaction (epistatic or environmental) , imperfect linkage disequilibrium (LD) with a causal variant, or haplotype effects. Identifying vQTL can therefore improve power to detect genetic interactions and aid in fine-mapping QTL.

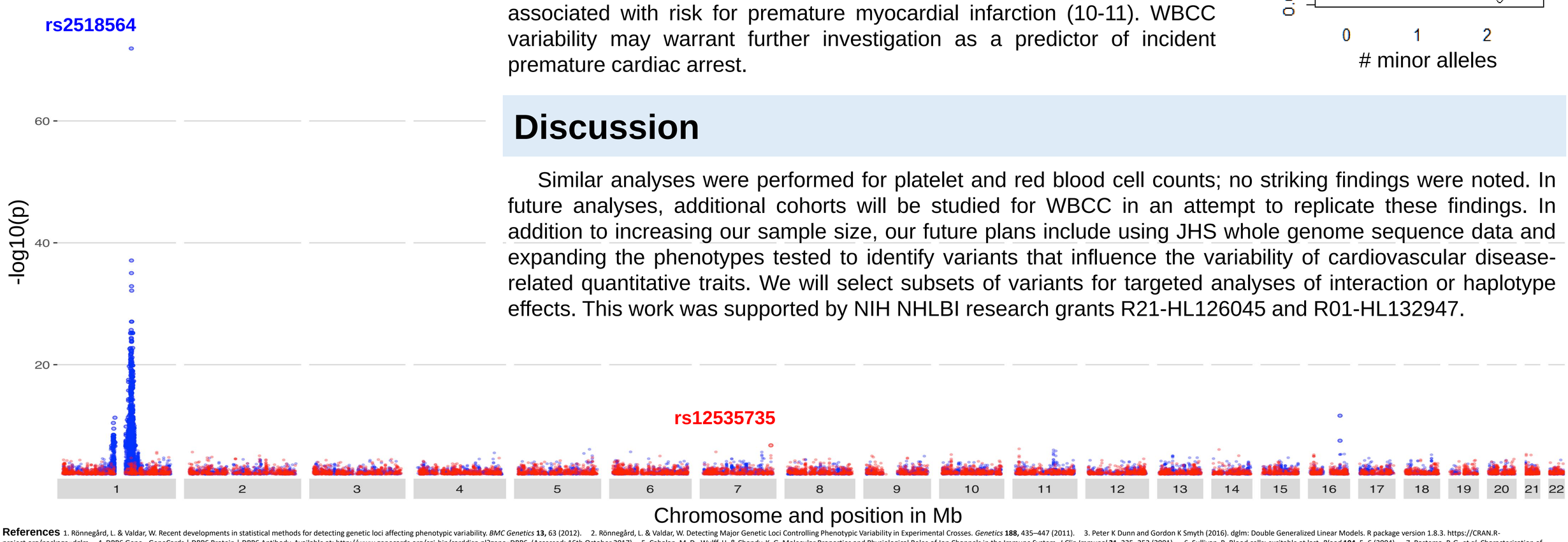
Figure 1. Variations on the additive model

a. Homoscedastic residual variance **b. Heteroscedastic residual variance**



The heteroscedastic model above (1b) depicts additive variance effects, however, vQTL effects may be non-additive.

Figure 2. Manhattan plot



Background (Continued)

The Double-Generalized Linear Model (DGLM)

The double-generalized linear model (DGLM) offers a full parametric, computationally efficient method for detecting variance effects with adjustment for mean effects. The DGLM includes a generalized linear model for the mean and a generalized linear sub-model for the residual variance. The residual variance model optionally allows the inclusion of fixed effects. Simultaneous estimation of mean and variance effects is achieved by iterating between models until convergence, with adjustment for error in the estimation of the residuals at each iteration (1-3). With genotype included as a fixed effect in the model for residual variance, a likelihood ratio test can be used to assess whether genotype is significantly associated with variability about the mean trait value.

Methods

We used the double generalized linear model (DGLM) to perform a genome scan for QTL and vQTL associated with WBCC in 2,257 unrelated African Americans from the Jackson Heart Study (JHS). Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Building on the R package DGLM, a separate DGLM was fit for each locus across the genome. WBCC was cleaned to exclude high outliers (WBCC>40) and normalized using a Box-Cox transformation ($\lambda = -0.22$). In both the mean and residual variance models, genotype was included as a predictor while age, sex, batch, and ten principal components for ancestry were included as covariates. Genomic control was applied to correct for inflation, after which our models showed good control of type I error.

Results

The top mean effect was observed at chromosome 1 variant rs2518564 ($p = 1.3 \times 10^{-72}$) in the previously established DARC region (Figure 1). Interestingly, this variant also showed nominal evidence for a variance effect ($p = 7.3 \times 10^{-3}$), suggesting either a genetic interaction (epistasis or environmental) or moderate linkage disequilibrium with the causal variant (likely rs2814778, a previously reported null variant in DARC that is monomorphic in non-African descent populations).

Table 1. Top mean and variance effects detected in genome-wide scan

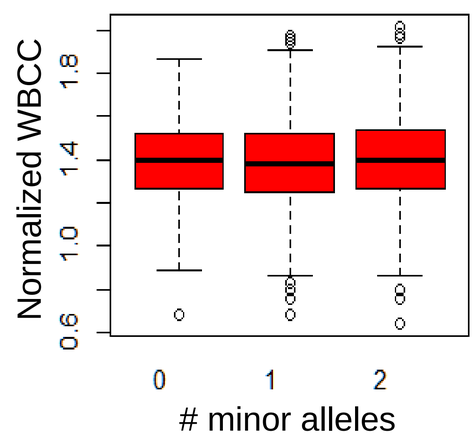
SNP	Chr.	Gene	Mean Effects			Variance Effects		
			β	(SE)	P-value	β	(SE)	P-value
rs2518564	1q23	DARC	-0.035	(0.0019)	1.28E-72	-0.194	(0.0690)	0.0073
rs12535735	7q36	DPP6	0.0004	(0.0014)	0.7939	0.248	(0.0452)	1.73E-07

Our top finding for a WBCC vQTL was at chromosome 7 variant rs12535735 ($p = 1.7 \times 10^{-7}$), an intronic variant in DPP6 with no evidence of a mean effect (p=0.79). Variants in DPP6 have been shown to alter voltage gated potassium channels, which coordinate electrical signaling for cardiac muscle contraction and may modulate immune function (4-6).

Haplotype effects within DPP6 have been associated with familial idiopathic ventricular fibrillation (IVF). IVF is an unpredictable condition, with premature cardiac arrest being the first and only symptom (7-9).

Elevated WBCC has been documented as an independent risk factor for cardiac arrest, while white blood cell telomere length has been associated with risk for premature myocardial infarction (10-11). WBCC variability may warrant further investigation as a predictor of incident premature cardiac arrest.

Figure 3. WBCC variability by rs12535735 genotype



Discussion

Similar analyses were performed for platelet and red blood cell counts; no striking findings were noted. In future analyses, additional cohorts will be studied for WBCC in an attempt to replicate these findings. In addition to increasing our sample size, our future plans include using JHS whole genome sequence data and expanding the phenotypes tested to identify variants that influence the variability of cardiovascular disease-related quantitative traits. We will select subsets of variants for targeted analyses of interaction or haplotype effects. This work was supported by NIH NHLBI research grants R21-HL126045 and R01-HL132947.

References 1. Rönnegård, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genetics* **13,** 63 (2012). 2. Rönnegård, L. & Valdar, W. Detecting Major Genetic Loci Controlling Phenotypic Variability in Experimental Crosses. *Genetics* **188,** 435–447 (2011). 3. Peter K Dunn and Gordon K Smyth (2016). dglm: Double Generalized Linear Models. R package version 1.8.3. https://CRAN.R-project.org/package=dglm 4. DPP6 Gene - GeneCards | DPP6 Protein | DPP6 Antibody. Available at: http://www.genecards.org/cgi-bin/carddisp.pl?gene=DPP6. (Accessed: 16th October 2017) 5. Cahalan, M. D., Wulff, H. & Chandy, K. G. Molecular Properties and Physiological Roles of Ion Channels in the Immune System. *J Clin Immunol* **21,** 235–252 (2001). 6. Sullivan, R. Blood cells: excitable at last. *Blood* **104,** 5–6 (2004). 7. Postema, P. G. *et al.* Characterisation of familial idiopathic ventricular fibrillation linked to DPP6. *Eur Heart J* **34,** (2013). 8. Alders, M. *et al.* Haplotype-Sharing Analysis Implicates Chromosome 7q36 Harboring DPP6 in Familial Idiopathic Ventricular Fibrillation. *Am J Hum Genet* **84,** 468–476 (2009). 9. Postema, P. G. The Quest for the Identification of Genetic Variants in Unexplained Cardiac Arrest and Idiopathic Ventricular Fibrillation. *PLOS Genetics* **9,** e1003480 (2013). 10. Haim, M., Boyko, V., Goldbourt, U., Battler, A. & Behar, S. Predictive value of elevated white blood cell count in patients with preexisting coronary heart disease: the Bezafibrate Infarction Prevention Study. *Arch. Intern. Med.* **164,** 433–439 (2004). 11. Brouilette, S., Singh, R. K., Thompson, J. R., Goodall, A. H. & Samani, N. J. White Cell Telomere Length and Risk of Premature Myocardial Infarction. *Arterioscleros, Thrombosis, and Vascular Biology* **23,** 842–846 (2003).