

## Washington University School of Medicine Digital Commons@Becker

---

### Open Access Publications

---

2010

# Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease

Scot J. Matkovich

*Washington University School of Medicine in St. Louis*

Derek J. Van Booven

*Washington University School of Medicine in St. Louis*

Anna Hindes

*Washington University School of Medicine in St. Louis*

Min Young Kang

*Washington University School of Medicine in St. Louis*

Todd E. Druley

*Washington University School of Medicine in St. Louis*

*See next page for additional authors*

Follow this and additional works at: [http://digitalcommons.wustl.edu/open\\_access\\_pubs](http://digitalcommons.wustl.edu/open_access_pubs)

---

### Recommended Citation

Matkovich, Scot J.; Van Booven, Derek J.; Hindes, Anna; Kang, Min Young; Druley, Todd E.; Vallania, Francesco L.M.; Mitra, Robi D.; Reilly, Muredach P.; Cappola, Thomas P.; and Dorn, Gerald W. II, "Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease." *The Journal of Clinical Investigation*. 120,1. 280-289. (2010).

[http://digitalcommons.wustl.edu/open\\_access\\_pubs/1625](http://digitalcommons.wustl.edu/open_access_pubs/1625)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [engeszer@wustl.edu](mailto:engeszer@wustl.edu).

---

**Authors**

Scot J. Matkovich, Derek J. Van Booven, Anna Hindes, Min Young Kang, Todd E. Druley, Francesco L.M. Vallania, Robi D. Mitra, Muredach P. Reilly, Thomas P. Cappola, and Gerald W. Dorn II



# Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing *HSPB7* polymorphisms associated with disease

Scot J. Matkovich,<sup>1</sup> Derek J. Van Booven,<sup>1</sup> Anna Hindes,<sup>1</sup> Min Young Kang,<sup>1</sup> Todd E. Druley,<sup>2,3</sup> Francesco L.M. Vallania,<sup>3</sup> Robi D. Mitra,<sup>3</sup> Muredach P. Reilly,<sup>4</sup> Thomas P. Cappola,<sup>4</sup> and Gerald W. Dorn II<sup>1</sup>

<sup>1</sup>Center for Pharmacogenomics, Department of Medicine, <sup>2</sup>Division of Pediatric Hematology and Oncology, Department of Pediatrics, and

<sup>3</sup>Center for Genome Sciences, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA.

<sup>4</sup>Penn Cardiovascular Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

**Sporadic heart failure is thought to have a genetic component, but the contributing genetic events are poorly defined. Here, we used ultra-high-throughput resequencing of pooled DNAs to identify SNPs in 4 biologically relevant cardiac signaling genes, and then examined the association between allelic variants and incidence of sporadic heart failure in 2 large Caucasian populations. Resequencing of DNA pools, each containing DNA from approximately 100 individuals, was rapid, accurate, and highly sensitive for identifying common and rare SNPs; it also had striking advantages in time and cost efficiencies over individual resequencing using conventional Sanger methods. In 2,606 individuals examined, we identified a total of 129 separate SNPs in the 4 cardiac signaling genes, including 23 nonsynonymous SNPs that we believe to be novel. Comparison of allele frequencies between 625 Caucasian nonaffected controls and 1,117 Caucasian individuals with systolic heart failure revealed 12 SNPs in the cardiovascular heat shock protein gene *HSPB7* with greater proportional representation in the systolic heart failure group; all 12 SNPs were confirmed in an independent replication study. These SNPs were found to be in tight linkage disequilibrium, likely reflecting a single genetic event, but none altered amino acid sequence. These results establish the power and applicability of pooled resequencing for comparative SNP association analysis of target subgenomes in large populations and identify an association between multiple *HSPB7* polymorphisms and heart failure.**

## Introduction

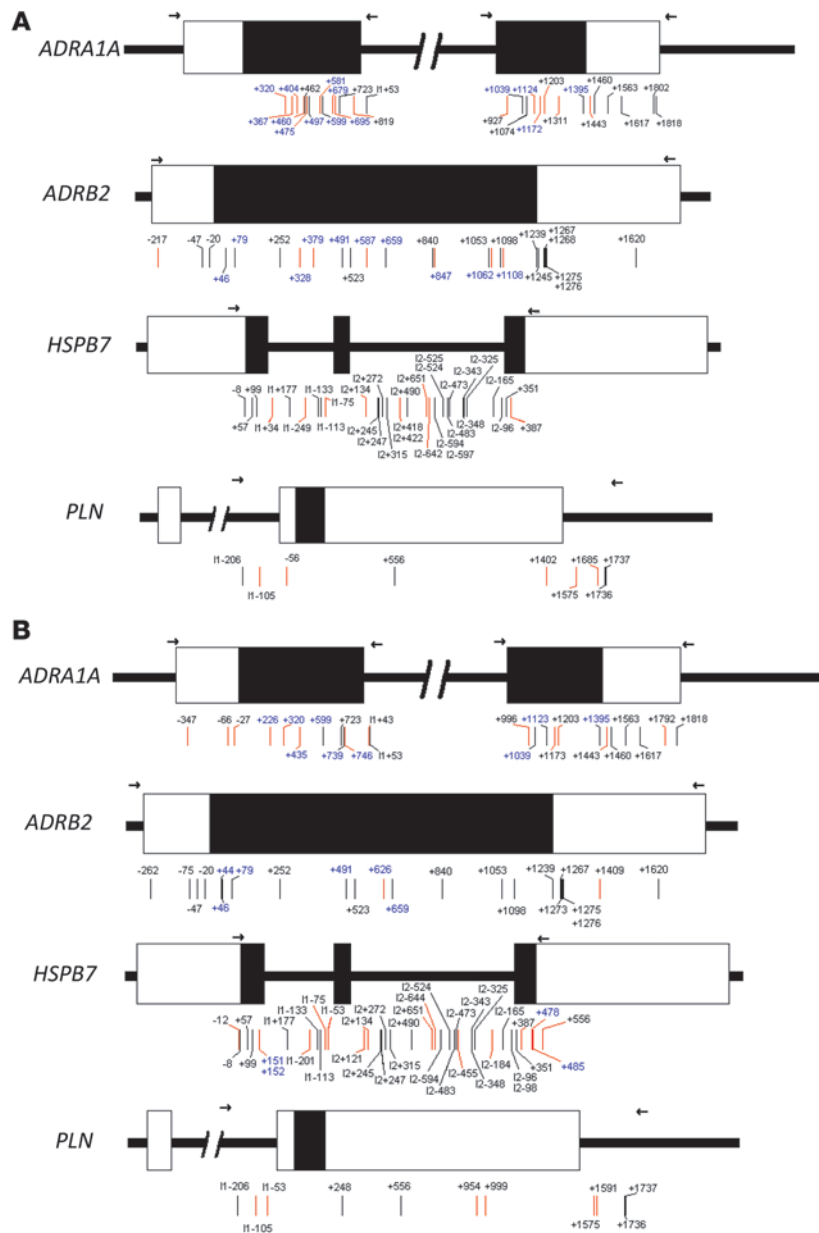
Application of genetic testing and linkage analysis to rare inherited disorders has uncovered thousands of causative genetic mutations and revealed much about disease pathophysiology and process (1). Identifying genetic causes for common diseases such as hypertension, diabetes, and heart failure has proven more challenging, likely because these complex syndromes develop as a consequence of interactions among multiple genetic and nongenetic factors. Furthermore, familial transmission of these disorders is inconsistent, disease penetrance tends to be incomplete, and the phenotypes can be highly variable. For these reasons, attempts to delineate genetic influences in common diseases have eschewed studies of familial transmission in favor of large-scale genetic association analyses that compare the frequencies of gene variants between affected cases and nonaffected controls. Genetic mapping of SNPs located at set intervals across the entire genome of thousands of individuals using microarray-based platforms has reproducibly implicated specific genetic loci in cancer, diabetes, hypertension, coronary atherosclerosis, and other diseases (2–6). However, assignment of the basis for gene effects and advancing our understanding into disease mechanisms requires resequencing across linked loci to identify causal mutations. In many cases, this has not yet been achieved.

Polygenic models for complex diseases propose that a significant component of genetic susceptibility is found in combinatorial interactions of multiple rare deleterious mutations (7, 8). Accordingly, the ultimate goal for studies correlating genotype and phenotype should be to identify all gene variants at multiple loci, both common and rare. This requires deep resequencing in very large and well-characterized patient populations (9). A first step is the NIH 1,000 Genomes Project, in which entire individual genomes are being independently resequenced at an estimated cost of at least \$30 million and will be collectively analyzed to generate a map of all human allelic variants with a frequency greater than 1% (10). More time- and cost-efficient gene mapping and identification of genetic modifiers for common diseases could be accomplished by targeted genetic analysis of pooled case and control samples. However, pooled genotyping is not compatible with conventional Sanger or microarray-based sequencing techniques.

We recently developed a targeted, pooled sample sequencing strategy (11) that, coupled with novel computational algorithms and second-generation DNA sequencing platforms (12, 13), accurately identified the position and frequency of common and rare SNPs in a random pediatric population of 1,111 individuals. Here, we have explored the applicability of these techniques to catalog nucleotide sequence diversity in a large adult population and to perform a SNP association study of systolic heart failure in 4 biologically relevant cardiac signaling genes:  $\alpha_1$ -adrenergic receptor

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Citation for this article:** *J. Clin. Invest.* 120:280–289 (2010). doi:10.1172/JCI39085.

**Figure 1**

Position of 129 signaling gene SNPs identified by pooled sequencing. Schematics show the exonic structure of the 4 target genes. Black boxes represent coding regions; white boxes represent 5' and 3' untranslated regions. Genomic PCR primer positions are shown with arrows. Detected SNPs are indicated by position and numbering from translation start site. Black lines indicate SNPs in dbSNP (19); red lines indicate novel SNPs. (A) Results for Caucasians (Cauc). (B) Results for African Americans (AA).

haplotype block of 12 synonymous or intronic HSPB7 SNPs in tight linkage disequilibrium that were significantly and reproducibly associated with systolic heart failure.

## Results

*Pooled DNA resequencing to detect and quantify SNPs in a large population.* To catalog nucleotide variations within the 4 target gene exons, we amplified and resequenced 5 loci covering 9.4 kb of nonamplified genomic DNA archived from 2,606 individuals — 1,742 Caucasians and 864 African Americans — who had participated in a longitudinal study of systolic heart failure (20). The coding regions for *ADRA1A*, *ADRB2*, *HSPB7*, and *PLN* were comprehensively examined (Figure 1). In initial studies determining the capabilities and limits of our pooled resequencing approach, genomic DNA from the study population was combined into 12 DNA pools of approximately 250 individuals per pool.

To determine the consistency of SNP calling by Illumina Genome Analyzer resequencing and our analysis pipeline, sequencing libraries were prepared from the 12 DNA pools and separately amplified and sequenced on 2 different machines at 2 different times. The data were analyzed in an identical manner, and allele frequencies were compared. The reported allele frequencies for confirmed SNPs were highly concordant between the 2 sequencing runs across the entire range

(*ADRA1A*),  $\beta_2$ -adrenergic receptor (*ADRB2*; refs. 14, 15), phospholamban (*PLN*; refs. 16, 17), and the cardiovascular heat shock protein (HSP) gene HSP27 member 7 (*HSPB7*; ref. 18). SNP calling using pooled sample resequencing was reproducible and, in comparison with individual Sanger sequencing of *ADRA1A* exon 2, both accurate and sensitive. We identified 55 common SNPs (*common* defined in the present study as greater than 1% overall allele frequency within a population of similar geographic ancestry) and 74 rare SNPs (*rare* defined as less than 1% overall allele frequency). Of these, 7 common SNPs were nonsynonymous (i.e., the genetic alteration changed an encoded amino acid), all of which are previously described in the SNP database (dbSNP; ref. 19), and 29 rare SNPs were nonsynonymous, 23 of which were novel. Comparison of common SNP allele frequencies in a Caucasian population (where *Caucasian* is defined as mixed European descent) between systolic heart failure cases and nonaffected controls revealed a

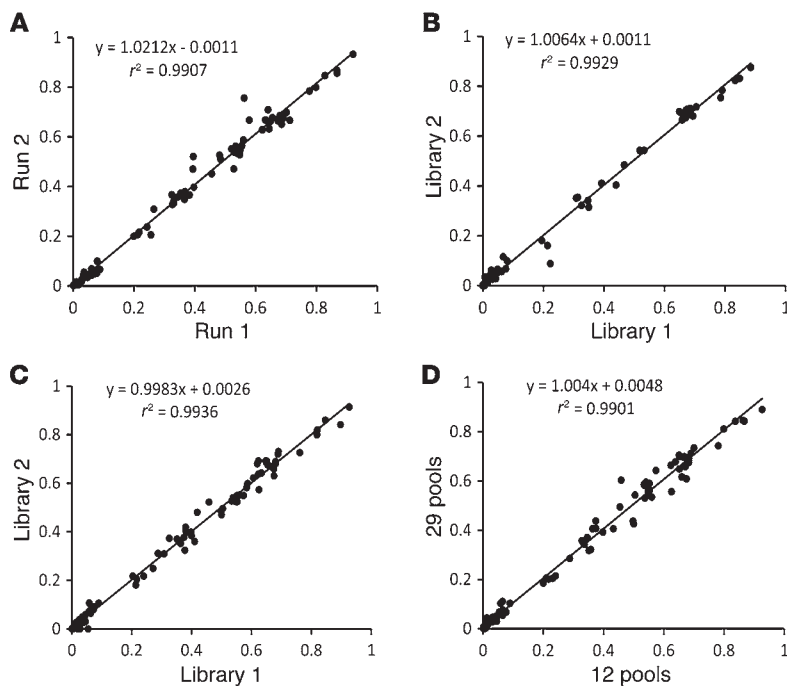
( $r^2 = 0.991$ ;  $P = 9 \times 10^{-122}$ ; Figure 2A).

Next, we evaluated the reproducibility of SNP identification and quantification as a function of preparing different sequencing libraries from the same DNA pools, or of starting with different DNA pools prepared from the same group of individual DNAs. This comparison examined variability introduced by differences in DNA handling and amplification. The correlation of reported allele frequencies in different sequencing libraries prepared from identical pools was as good as that from libraries prepared from different pools containing the same collection of individual DNAs ( $r^2 = 0.993$  and  $0.994$ , respectively;  $P < 1 \times 10^{-70}$ ; Figure 2, B and C).

Depth of sequence coverage using the pooled DNA sequencing approach is determined in part by the number of individual genomes added per single Illumina lane; for our experimental design, this corresponded with the number of individuals in each pool. For our initial studies using 12 pools of about 250



## research article

**Figure 2**

Allele frequency comparisons with Illumina resequencing of pooled DNA. **(A)** Comparison of SNP allele frequencies from 2 independent runs of identical sequencing libraries made from 12 pools of 2,606 individual DNAs. **(B)** Comparison of SNP allele frequencies from 2 independent runs of separate libraries made from the same DNA pools. **(C)** Comparison of SNP allele frequencies from 2 independent runs of libraries made from replicate DNA pools. **(D)** Comparison of SNP allele frequencies obtained from resequencing the same 2,606 individual DNAs, grouped into pools of approximately 250 individual DNAs (12 pools) versus approximately 100 individual DNAs (29 pools). Formulae for regression line and Pearson correlation coefficient are indicated for each analysis.

individuals, sequence coverage for SNP calling (first 12 nucleotides only) averaged 10-fold per diploid genome. Coverage depth for the noncoding region of *PLN*, which has low sequence complexity due to the presence of multiple dinucleotide repeats, was not as great (3-fold per diploid genome). To assess the impact of increasing sequence coverage on SNP calling, we resequenced the 5 genetic loci in the study cohort divided into 29 pools containing approximately 100 individual DNAs (range, 75 to 125). With this DNA pooling design and intercurrent improvements in Illumina sequencing chemistry (21), and by sequencing the library from each pool in a separate lane, we found average sequence coverage depth increased to 43-fold (11-fold for *PLN* noncoding).

SNP identification was compared between the 250 individuals per pool (12-pool) and 100 individuals per pool (29-pool) studies. A total of 129 separate SNPs (55 common, 74 rare) were identified and validated in the 29-pool study. Correlation of allele frequencies for SNPs reported in both the 12- and 29-pool studies was excellent ( $r^2 = 0.990$ ,  $P = 4.2 \times 10^{-115}$ ; Figure 2D). Of 81 SNPs reported in the 12-pool study, 77 (95%) were detected in the 29-pool study, but 49 validated SNPs reported in the 29-pool design had not fulfilled validation criteria in the 12-pool studies. Of these newly validated SNPs, 47 were rare (Caucasians, median allele frequency 0.0008, range 0.0003 to 0.005; African Americans, median allele frequency 0.014, range 0.0006 to 0.019), and 2 were private (Caucasians, allele frequency 0.0003; African Americans, allele frequency 0.0006). Thus, DNA pooling into smaller groups of about 100 individuals improved detection of rare SNPs by increasing sequence coverage depth and by affording greater opportunity for SNP calling in multiple different pools.

We examined the sensitivity and specificity of Illumina pooled resequencing in comparison to the previous gold standard, conventional Sanger resequencing of individual DNAs. Pooled sequencing showed that the 5' end of the *ADRA1A* exon 2 contains 15 SNPs (7 reported in HapMap and 8 novel), with allele frequencies ranging from approximately 50% to a singleton (Figure 1 and Table 1).

Therefore, we individually Sanger resequenced this amplicon for each subject DNA. Of the 2,606 DNAs amplified and submitted, confident Sanger sequence was returned on 2,506 (1,666 Caucasians and 840 African Americans; 96% overall sequence success rate). SNP allele frequencies generated by pooled and individual resequencing are compared in Table 1, reported separately for Caucasians and African Americans. Pooled Illumina resequencing and individual Sanger resequencing identified the same 15 SNPs. Pooled sequencing detected 5 confirmed singletons in the Caucasian subgroup (+1,074, +1,172, +1,203, +1,311, and +1,443) and 1 SNP that occurred only twice in this group (+927). However, within the Caucasian and African American subgroups, pooled resequencing failed to detect a rare SNP in the Caucasian subgroup that was detected on Sanger sequence (+996); this also occurred once in the African American cohort (+1,124). In both instances, the SNP had been bioinformatically filtered from the pooled sequencing results because it was novel and not detected in more than 1 DNA pool, or in both the 29-pool and 12-pool studies. For the 3 common SNPs (+1,039, +1,395, and +1,617), the allele frequencies generated by pooled Illumina sequencing were comparable to those from individual Sanger sequencing. These results show that allele frequencies for common SNPs obtained by Illumina resequencing of pools of about 100 individual DNAs were sufficiently accurate for between-group comparisons. These data also demonstrate that pooled resequencing and our analysis pipeline can be highly sensitive for the detection of rare and private SNPs in large populations, although stochastic variation and our filtering criteria made for less than absolute sensitivity in all pools. Based on this finding, we determined that quantitative comparisons of individual SNP allele frequencies between population groups are best performed on common SNPs and that a more qualitative approach should be used to evaluate rare and private SNPs.

*Allele frequencies in population subgroups with different geographic ancestries.* Resequencing target genes across a large population should be useful to identify alleles that are differentially represented between



**Table 1**  
*ADRA1A* exon 2 SNP allele frequencies from pooled Illumina versus individual Sanger sequencing

dbSNP	Position	Type	Major allele	Minor allele	Codon	Pooled Illumina		Individual Sanger		
						Cauc	AA	Cauc	AA	
rs1048101	+927	Coding S	A	G	K309K	0.0004	–	0.0006	–	
	+996	Coding S	G	A	E332E	–	0.0014	0.0009	0.0027	
	+1,039	Coding NS	T	C	C347R	0.4373	0.7435	0.4856	0.7611	
	rs56318220	+1,074	Coding S	C	T	Y358Y	0.0003	–	0.0004	–
	rs61731555	+1,123	Coding NS	G	T	V375L	–	0.0017	–	0.0016
	+1,124	Coding NS	T	G	V375G	0.0006	–	0.0015	0.0024	
	+1,172	Coding NS	C	T	T391M	0.0003	–	0.0004	–	
	+1,173	Coding S	G	A	T391T	–	0.0023	–	0.0031	
	+1,203	Coding S	T	G	S401S	0.0004	0.0019	0.0004	0.0031	
	+1,311	Coding S	G	T	G437G	0.0003	–	0.0004	–	
rs2229126	+1,395	Coding NS	A	T	E465D	0.0308	0.0278	0.0232	0.0221	
	+1,443	3'UTR	C	A		0.0003	0.0008	0.0004	–	
rs61760513	+1,460	3'UTR	C	A		0.0003	0.0062	0.0009	0.0043	
rs61760511	+1,563	3'UTR	G	T		0.0035	0.0013	0.0048	0.0035	
rs73678233	+1,617	3'UTR	C	T		0.0012	0.0661	0.0015	0.0487	

Illumina sequencing was performed on 19 pools of 1,742 Caucasians (Cauc) and 9 pools of 864 African Americans (AA). Sanger sequencing was successful for 1,666 Caucasians and 840 African Americans assayed individually. For SNPs in HapMap, the rs number from dbSNP (19) is shown. Position is from the translation initiation codon. S, synonymous; NS, nonsynonymous; UTR, untranslated region. Values in the codon column show the amino acid. Allele frequencies of ~0.0003 and ~0.0006 correspond to singletons in Caucasians and African Americans, respectively.

population subgroups. We recently showed that differences in SNP frequency between 2 groups with different genetic ancestries can confound the results of systolic heart failure outcome studies when genotype is not an a priori consideration in experimental design (20). Racial differences in SNP frequencies are reported in the HapMap, but the number of individuals from some ancestral subgroups can be small. Therefore, we used pooled DNA resequencing to examine SNP diversity within our 4 target genes in 625 Caucasian and 236 African American subjects who did not have heart disease at the time of study entry. A total of 110 SNPs were detected in the normal subjects, 48 of which were novel (Table 2 and Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI39085DS1). We found 27 SNPs to be nonsynonymous (encoded a change in amino acid), 16 of which were novel (Table 2 and Supplemental Table 1). Even though our sample number for African Americans was only about one-third that for Caucasians, African Americans had 84 SNPs compared with 69 in Caucasians. Moreover, 41 SNPs were exclusive to African Americans (including 7 in the 3' untranslated region of *ADRB2* and 15 in the second intron and third exon of *HSPB7*), and 26 SNPs were exclusive to Caucasians (including 12 in the coding region of *ADRA1A* and 7 in the coding region of *ADRB2*; Supplemental Table 1 and Figure 3A). Of the 42 SNPs present in both population subgroups, the allele frequencies differed significantly ( $P < 0.0005$ ;  $\alpha = 0.05$ ) between African Americans and Caucasians in 27, and in 24 of these instances, the allele frequency was significantly greater in African Americans (Table 2 and Figure 3A). These results reveal substantial nucleotide sequence diversity in the 4 target genes as a function of geographic ancestry and support SNP association study designs that analyze Caucasians and African Americans separately.

*Common signaling gene polymorphisms associated with systolic heart failure.* Another instance in which pooled resequencing should have advantages over individual resequencing is to identify alleles (espe-

cially those that are not represented on marker SNP arrays) that are differentially represented between diseased cases and nonaffected controls. Accordingly, we used pooled DNA resequencing for a case-control gene association study of systolic heart failure among Caucasians. We compared allele frequencies for the 37 common SNPs identified in our pooled DNA resequencing of 1,117 Caucasian heart failure cases (the 29-pool study) with 625 Caucasian controls from the University of Cincinnati, and confirmed associated SNPs in a replication pooled sequencing analysis of 859 Caucasian heart failure cases and 311 Caucasian controls from the University of Pennsylvania. Detailed clinical characteristics of the systolic heart failure study population, stratified by geographic ancestry, have been reported elsewhere (20). A summary of demographic and clinical features of the study subjects is shown in Table 3. The DNA pooling strategy

for the study groups is shown in Supplemental Figure 1.

Our 4 target genes were selected because of their postulated relevance to cardiac physiology and disease (22–24). Therefore, we were gratified to find genotype-phenotype associations (Table 4 and Supplemental Table 2). Of the 37 common SNPs analyzed in the case-control cohort, 12 SNPs spanning exons 1–3 of *HSPB7*, including in the intervening 2 introns, achieved the  $P < 0.0014$  ( $\alpha = 0.05$ ) threshold for significance in the primary cohort (Figure 3B and Table 4). These SNPs were in tight linkage disequilibrium (Figure 3B), and therefore likely indicate a common genetic event. The association of each of these *HSPB7* SNPs was replicated in a subsequent study of independent heart failure cases and controls ( $P < 0.0211$ ;  $\alpha = 0.05$ ; effective number of independent SNPs, 2.36; Table 4).

The heart failure study cohorts consisted of a mixture of ischemic and nonischemic (also known as idiopathic) cardiomyopathies, but our DNA pooling strategy permitted us to separately examine the association of *HSPB7* polymorphisms with each etiology. In the primary cohort of 691 with ischemic cardiomyopathy and 426 with nonischemic cardiomyopathy, multiple SNPs within the *HSPB7* haplotype block were significant in both etiological subgroups (Supplemental Table 3), which suggests that the association of this *HSPB7* haplotype applies to both ischemic and nonischemic cardiomyopathies.

Although we were underpowered to perform a comprehensive SNP association analysis of African American systolic heart failure, we examined our limited cohort of 628 African American cases and 236 African American controls to determine whether any of the *HSPB7* SNP associations detected in the primary Caucasian cohort were present in this subgroup. African American systolic heart failure was similar to Caucasian heart failure in terms of age, height, weight, and degree of left ventricular dysfunction (ejection fraction), but, consistent with previous observations (25, 26), hypertension and nonischemic cardiomyopathy were more prevalent (20).



## research article

**Table 2**  
Comparative allele frequencies of target gene SNPs in Caucasians and African Americans

dbSNP	Position	Type	Major allele	Minor allele	Codon	Cauc controls	AA controls	P
<b>ADRA1A</b>								
	+320	Coding NS	T	G	V17G	0.0024	0.0079	0.0945
rs2229125	+599	Coding NS	T	G	I200S	0.0216	0.0017	0.0022
rs382423	I1+53	Intronic	G	T		0.0162	0.0095	0.3555
rs14811	+1,039	Coding NS	T	C	C347R	0.4317	0.7023	<0.0001 <sup>A</sup>
rs2229126	+1,395	Coding NS	A	T	E465D	0.0243	0.0226	1.0000
rs6176513	+1,460	3'UTR	C	A		0.0008	0.0110	0.0071
rs6176511	+1,563	3'UTR	G	T		0.0030	0.0021	1.0000
rs73678233	+1,617	3'UTR	C	T		0.0011	0.0591	<0.0001 <sup>A</sup>
rs3739216	+1,818	3'UTR	C	G		0.0742	0.1800	<0.0001 <sup>A</sup>
<b>ADRB2</b>								
rs142711	-47	5'UTR	C	T		0.6161	0.8451	<0.0001 <sup>A</sup>
rs18174	-20	5'UTR	C	T		0.6185	0.8415	<0.0001 <sup>A</sup>
rs142713	+46	Coding NS	G	A	G16R	0.3985	0.4781	0.0031
rs142714	+79	Coding NS	G	C	E27Q	0.5716	0.8165	<0.0001 <sup>A</sup>
rs142717	+252	Coding S	G	A	L84L	0.2165	0.3965	<0.0001 <sup>A</sup>
rs142718	+523	Coding S	C	A	R175R	0.2004	0.3892	<0.0001 <sup>A</sup>
rs142719	+1,053	Coding S	G	C	G351G	0.2995	0.3851	0.0008 <sup>A</sup>
rs14272	+1,239	Coding S	G	A	L413L	0.3256	0.5402	<0.0001 <sup>A</sup>
rs8192451	+1,620	3'UTR	C	T		0.0150	0.0064	0.2267
<b>HSPB7</b>								
rs373864	-8	5'UTR	G	A		0.1061	0.0410	<0.0001 <sup>A</sup>
rs945416	+57	Coding S	C	T	S19S	0.5383	0.6737	<0.0001 <sup>A</sup>
rs732286	+99	Coding S	T	C	A33A	0.5171	0.6639	<0.0001 <sup>A</sup>
rs3883891	I1+177	Intronic	A	G		0.3730	0.3590	0.5762
rs1739845	I1-113	Intronic	G	T		0.0501	0.0788	0.0369
rs3738643	I2+245	Intronic	C	G		0.1116	0.0404	<0.0001 <sup>A</sup>
rs1739844	I2+247	Intronic	A	G		0.5109	0.6735	<0.0001 <sup>A</sup>
rs1763597	I2+272	Intronic	G	A		0.5276	0.6988	<0.0001 <sup>A</sup>
rs1739843	I2+315	Intronic	A	G		0.5432	0.7014	<0.0001 <sup>A</sup>
rs12758813	I2+490	Intronic	C	T		0.4173	0.4350	0.5476
rs1739842	I2-594	Intronic	A	C		0.5235	0.6975	<0.0001 <sup>A</sup>
rs1763598	I2-483	Intronic	A	G		0.8775	0.8424	0.0662
rs1739841	I2-473	Intronic	T	C		0.4802	0.6451	<0.0001 <sup>A</sup>
rs1763599	I2-348	Intronic	G	A		0.6009	0.7442	<0.0001 <sup>A</sup>
rs61783988	I2-343	Intronic	T	C		0.0779	0.0305	0.0002 <sup>A</sup>
rs76176	I2-165	Intronic	T	C		0.5307	0.6803	<0.0001 <sup>A</sup>
rs761759	I2-96	Intronic	T	A		0.5519	0.7080	<0.0001 <sup>A</sup>
rs173984	+351	Coding S	C	T	T117T	0.4735	0.3096	<0.0001 <sup>A</sup>
<b>PLN</b>								
rs73526175	I1-206	Intronic	G	T		0.0030	0.0710	<0.0001 <sup>A</sup>
	I1-105	Intronic	A	G		0.0013	0.0043	0.3029
rs12198461	+556	3'UTR	T	G		0.5238	0.4181	<0.0001 <sup>A</sup>
	+1,575	3'FR	G	A		0.0209	0.0624	<0.0001 <sup>A</sup>
rs9489441	+1,736	3'FR	G	C		0.0029	0.0781	<0.0001 <sup>A</sup>
rs151429	+1,737	3'FR	C	T		0.7017	0.7003	0.9529

Allele frequencies from Illumina resequencing of 7 DNA pools from 625 nonaffected Caucasians and 3 DNA pools from 237 nonaffected African Americans. For SNPs in HapMap, the rs number from dbSNP (19) is shown. Position of exonic SNPs is relative to +1 position of the translation initiation codon, and of SNPs within introns is relative to the nearest exon. I1, first intron; I2, second intron; S, synonymous; NS, nonsynonymous; UTR, untranslated region; FR, flanking region. Values in the codon column show the amino acid. Only those SNPs found to occur in both Caucasian and African Americans are shown; Supplemental Table 1 shows all 110 SNPs. <sup>A</sup>P < 0.0005, the threshold for this study ( $\alpha = 0.05$ ,  $n = 110$ ).

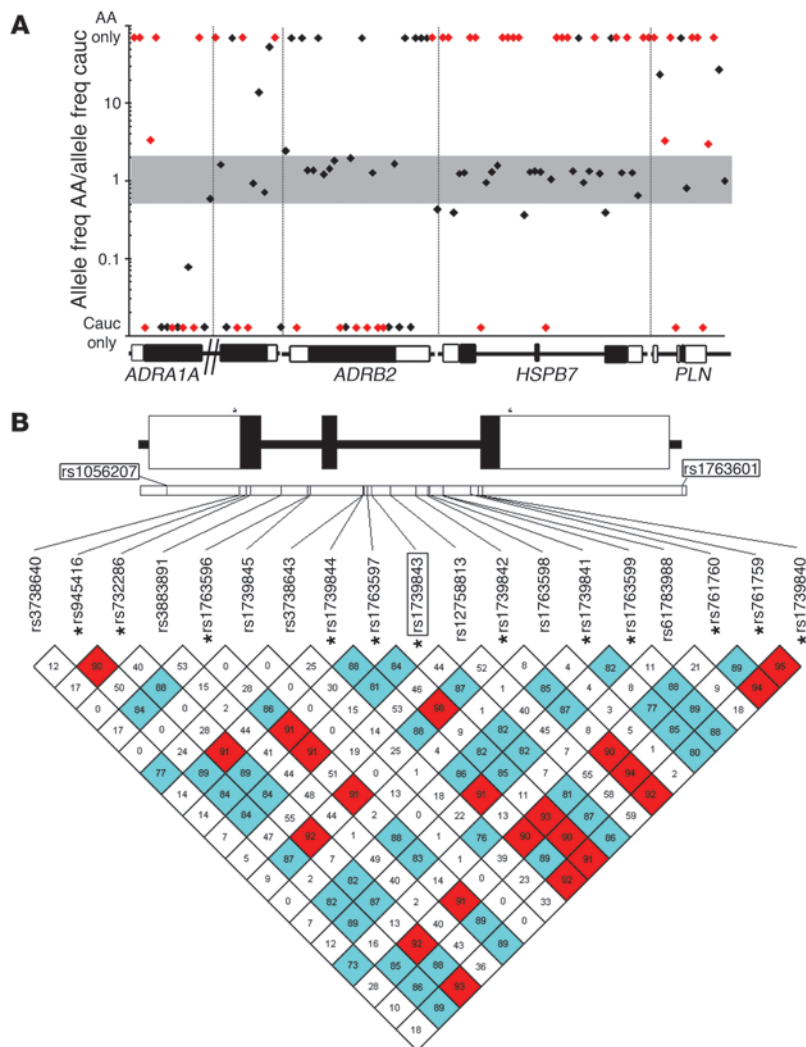
Caucasians examined only common SNPs. Although 3 of the 12 associated *HSPB7* SNPs were in coding exons, they did not alter amino acid sequence (Figure 3B and Table 4). Nevertheless, it is estimated that 50% of rare nonsynonymous SNPs can be deleterious (27). For this reason, we investigated the impact on heart failure by rare and common nonsynonymous SNPs in the 4 target genes. First, we used computational algorithms SIFT and PolyPhen (28, 29) to categorize the likelihood that the nonsynonymous SNPs could be harmful, based upon sequence conservation, protein structure, and annotated protein features. All coding SNPs in the target genes were classified as either synonymous (expected to be biologically neutral), nonsynonymous and presumed benign (both algorithms agreed), nonsynonymous and presumed pathological (both algorithms agreed), or nonsynonymous with uncertain consequences (algorithms differed) (Supplemental Tables 5 and 6). Case-control comparisons of allele frequencies were performed for common SNPs (analyzed individually; Supplemental Table 5) and rare SNPs (analyzed individually and collectively as functionally similar groups; Supplemental Table 6). For common SNPs, there were 7 nonsynonymous SNPs, only one of which (*ADRA1A* I200S) was classified as presumed pathological. Allele frequencies for this SNP were approximately 0.02 in Caucasians and approximately 0.002 in African Americans, with no differences between heart failure cases and controls. For rare SNPs, there were 29 nonsynonymous SNPs (of which 16 were in *ADRA1A*): 13 were classified as presumed benign, 10 as presumed pathological, and 6 as uncertain (Supplemental Table 6). Again, there were no differences in allele frequencies between heart failure cases and controls for these SNPs, either individually or combined into functionally similar groups.

There was no association among any of the *HSPB7* SNPs with systolic heart failure in African Americans (Supplemental Table 4).

*Impact of nonsynonymous signaling gene polymorphisms on systolic heart failure.* Our case-control study of systolic heart failure in

## Discussion

Here, we used large-scale pooled sequencing of phenotypically similar individuals to identify and quantify common and rare SNPs in a target cardiac signaling subgenome and applied this

**Figure 3**

Comparative analysis of SNP allele frequencies in population subgroups. **(A)** Comparison of target gene SNP allele frequencies in African Americans versus Caucasians. Vertical axis is African American allele frequency/Caucasian allele frequency, plotted on a  $\log_{10}$  scale. Horizontal axis shows target gene with SNPs positioned accordingly, as in Figure 1. Red symbols denote novel SNPs. The shaded area indicates less than 2-fold difference in relative allele frequency. Symbols at the extreme top and bottom represent alleles seen exclusively in African Americans and Caucasians, respectively (see Table 2). **(B)** Linkage analysis of *HSPB7* SNPs in Caucasians and association with systolic heart failure. *HSPB7* SNP positions are shown within the gene (schematic as in Figure 1; boxes denote SNPs present in HapMap). Allele frequencies were compared across pools of 100 individuals. Red shading in the linkage plot denotes  $r^2 \geq 0.90$ ; blue shading denotes  $r^2 > 0.70$ ; no shading denotes  $r^2 < 0.70$ . Numbers within plot show  $r^2$  values ( $\times 100$ ). Asterisks denote SNPs significantly associated with Caucasian heart failure ( $P < 0.0014$ ; see Table 4).

information to a case-control SNP association analysis of sporadic ischemic and nonischemic cardiomyopathy. To our knowledge, this is the first large-scale application of next-generation sequencing technologies to study a disease using pooled DNA samples; therefore, we were careful to validate the approach and to replicate our findings, both within the primary study cohort and in an independent systolic heart failure cohort. We found that Illumina resequencing of DNA pools containing about 100 individual DNAs was rapid, accurate, and highly sensitive for identifying common and rare or private SNPs and had striking advantages in time and cost efficiencies over individual resequencing. Our results identified, and verified by independent replication, a cluster of SNPs within a haplotype block of *HSPB7* that were associated with development of systolic heart failure. Interestingly, none of the associated SNPs altered amino acid coding of *HSPB7* protein, nor were nonsynonymous SNPs independently associated with heart failure, even those predicted to be functionally deleterious. The present studies demonstrate the power and potential broad applicability of pooled resequencing for comparative SNP association analysis of target subgenomes in large populations and describe what we believe to be a previously unsuspected association between *HSPB7* and sporadic heart failure.

*Applicability of pooled resequencing with next-generation platforms.* Much of the low-hanging fruit from human genetic studies has been picked over. Thousands of disease-causing gene mutations have been identified, and millions of relatively common gene polymorphisms have been cataloged, yet the impact of these findings has primarily been on rare inherited diseases that are almost entirely genetic in cause. The next step is to elucidate the more modest genetic effects attributable to simultaneous influences of multiple noncausal gene variants on common diseases that are not primarily genetic (8). This requires sequencing many thousands of individual human genomes at a resolution of individual nucleotides, which we accomplished via a recently developed and validated pooled sequencing strategy (11). Our procedure accurately detected variant alleles across a wide range of frequencies in Caucasian and African American normal controls and in heart failure subjects. Results for *ADRA1A* exon 2, in which there were SNPs covering the entire range of allele frequencies (but no difference in allele frequencies between heart failure cases and controls), were confirmed by individual Sanger sequencing of the entire study cohort. Importantly, this pooled sequencing protocol generated these data for a much smaller investment of time and resources compared with conventional methods that rely upon serial geno-





## research article

**Table 3**

Clinical characteristics of study cohorts

Variable	Primary controls		Primary CHF		Secondary controls		Secondary CHF	
	Cauc	AA	Cauc	AA	Cauc	Cauc	AA	AA
<i>N</i>	625	236	1,117	628	311			859
Age (yr)	48.74 ± 12.0	54.0 ± 13.3	57.1 ± 13.2	54.0 ± 13.3	58.0 ± 10.9			58.5 ± 13.1
Height (in.)	67.6 ± 4.2	66.4 ± 5.01	67.7 ± 4.1	67.0 ± 4.2	67.7 ± 4.0			68.2 ± 4.0
Weight (lb)	182.7 ± 45.1	194.3 ± 46.5	193.0 ± 50.0	197.2 ± 55.9	186.4 ± 43.7			189.2 ± 44.5
Ejection fraction (%)	–	–	28.7 ± 13.2	26.9 ± 11.5	66.0 ± 6.3			31.8 ± 16.2
CHF etiology ( <i>n</i> )								
Ischemic	–	–	689 (60%)	295 (47%)	–			286 (50%)
Nonischemic	–	–	427 (40%)	332 (53%)	–			285 (50%)
Female	375 (60%)	184 (77%)	359 (32%)	285 (46%)	138 (44%)			158 (28%)
Hypertension	16 (3%)	1 (0%)	794 (69%)	568 (91%)	144 (47%)			252 (44%)

typing of each individual in each cohort. Extrapolating the time and cost of individual capillary resequencing versus Illumina resequencing of pools of about 100 individual DNAs (referred to herein as the 29-pool study), accurate allele frequencies and detection of very rare and private SNPs was achieved in the current instance for one-sixteenth the cost and about one-thirtieth the time and effort. If the experimental design requires only quantification of alleles with frequencies greater than 0.01 (i.e., prevalence more than 1%), then there appears to be no disadvantage to using a larger, about 250-individual DNA pooling strategy at an additional doubling of the cost savings. These cost and time efficiencies are expected to further improve with double-end reading and further

technological advances, and the protocols can be applied to any genes in virtually any population of interest. For example, pooled resequencing of large, existing cardiovascular study populations could rapidly identify and quantify biologically relevant SNPs in cardiac contractile protein genes, ion pump and channel genes, or transcription factor genes (30–32).

*SNP associations with systolic heart failure.* The current study focused on 4 signaling genes, selected for their potential importance in cardiac disease and (in these initial pooled resequencing studies) for their uncomplicated genetic structure. Polymorphisms of the  $\alpha_1$ -adrenergic receptor, encoded by *ADRA1A*, have been linked with hypertensive disease in some, but not all, populations (33–35). It

**Table 4**Comparative allele frequencies of common *HSPB7* SNPs in Caucasian heart failure

dbSNP	Position	Type	Major allele	Minor allele	Codon	Primary controls	Primary cases	Fisher exact	Adjusted $\chi^2$	Secondary controls	Secondary cases	Fisher exact	Adjusted $\chi^2$
rs3738640	–8	5'UTR	G	A		0.106	0.102	0.6847	0.7246				
rs945416	+57	Coding S	C	T	S19S	0.538	0.606	0.0001 <sup>A</sup>	0.0001 <sup>A</sup>	0.510	0.582	0.0025 <sup>B</sup>	0.0001 <sup>B</sup>
rs732286	+99	Coding S	T	C	A33A	0.517	0.594	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.489	0.575	0.0003 <sup>B</sup>	<0.0001 <sup>B</sup>
rs3883891	I1+177	Intronic	A	G		0.373	0.424	0.0032	0.0040				
rs1763596	I1-133	Intronic	G	A		0.538	0.619	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.506	0.576	0.003 <sup>B</sup>	0.0001 <sup>B</sup>
rs1739845	I1-113	Intronic	G	T		0.050	0.052	0.8732	0.8168				
rs3738643	I2+245	Intronic	C	G		0.112	0.098	0.2454	0.2156				
rs1739844	I2+247	Intronic	A	G		0.511	0.589	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.480	0.571	0.0001 <sup>B</sup>	<0.0001 <sup>B</sup>
rs1763597	I2+272	Intronic	G	A		0.528	0.618	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.515	0.575	0.0109 <sup>B</sup>	0.0008 <sup>B</sup>
rs1739843	I2+315	Intronic	A	G		0.543	0.624	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.495	0.585	0.0001 <sup>B</sup>	<0.0001 <sup>B</sup>
rs12758813	I2+490	Intronic	C	T		0.417	0.450	0.0696	0.0654				
rs1739842	I2-594	Intronic	A	C		0.524	0.615	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.480	0.568	0.0002 <sup>B</sup>	<0.0001 <sup>B</sup>
rs1763598	I2-483	Intronic	A	G		0.878	0.898	0.0796	0.0836				
rs1739841	I2-473	Intronic	T	C		0.480	0.578	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.405	0.506	<0.0001 <sup>B</sup>	<0.0001 <sup>B</sup>
rs1763599	I2-348	Intronic	G	A		0.601	0.666	0.0001 <sup>A</sup>	0.0002 <sup>A</sup>	0.554	0.633	0.0006 <sup>B</sup>	<0.0001 <sup>B</sup>
rs61783988	I2-343	Intronic	T	C		0.078	0.075	0.7903	0.7652				
rs761760	I2-165	Intronic	T	C		0.531	0.598	0.0001 <sup>A</sup>	0.0002 <sup>A</sup>	0.484	0.570	0.0003 <sup>B</sup>	<0.0001 <sup>B</sup>
rs761759	I2-96	Intronic	T	A		0.552	0.632	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.456	0.581	<0.0001 <sup>B</sup>	<0.0001 <sup>B</sup>
rs1739840	+351	Coding S	C	T	T117T	0.473	0.399	<0.0001 <sup>A</sup>	<0.0001 <sup>A</sup>	0.561	0.443	<0.0001 <sup>B</sup>	<0.0001 <sup>B</sup>

Allele frequencies of common SNPs in *HSPB7* from Illumina resequencing of 7 DNA pools from 625 nonaffected primary controls, 12 DNA pools from 1,117 primary cases, 4 DNA pools from 311 nonaffected secondary controls, and 10 DNA pools from 825 secondary cases. For SNPs in HapMap, the rs number from dbSNP (19) is shown. Position is from the translation initiation codon. I1, first intron; I2, second intron; S, synonymous; UTR, untranslated region. Values in the codon column show the amino acid. Comparative allele frequencies of all 37 common target gene SNPs are given in Supplemental Table 2. Statistical comparisons used Fisher exact and adjusted  $\chi^2$  tests as described in Methods. <sup>A</sup> $P < 0.0014$ , the threshold for the primary study ( $\alpha = 0.05$ , 2-tailed test;  $n = 37$ ). <sup>B</sup> $P < 0.0212$ , the threshold for the replication study ( $\alpha = 0.05$ , 2-tailed test; effective number of independent SNPs, 2.36).



is striking that our studies detected 14 nonsynonymous polymorphisms of this receptor, 8 of which were rare and not previously reported in dbSNP (Table 2 and ref. 19). To our knowledge, there are no published reports of associations between *ADRA1A* variants and development of heart failure, and no significant associations were detected in our studies. Likewise,  $\beta_2$ -adrenergic receptors, encoded by *ADRB2*, mediate the cardiac inotropic response to catecholamines and are the targets of  $\beta$ -blocking drugs. This gene is highly polymorphic (10 nonsynonymous SNPs in our study, 5 rare and previously unreported), but does not appear to contribute to heart failure risk in our present study or in previous studies (36). Phospholamban, encoded by *PLN*, is a critical regulator of cardiomyocyte calcium cycling that transmits inputs from the  $\beta$ -adrenergic receptor system to the major sarcoplasmic reticulum calcium reuptake pump, SERCA2 (37). In dramatic contrast to the results for the 2 adrenergic receptors, not a single nonsynonymous polymorphism for *PLN* was detected in the more than 2,600 individuals of our cohort. This remarkable degree of amino acid sequence conservation was consistent with the severe pathological consequences previously described with loss-of-function *PLN* mutants. Substitution of cysteine for arginine at amino acid 9 (16), introduction of a premature stop codon at amino acid 39 (38), and deletion of arginine 14 (17, 39) each cause a familial cardiomyopathy with potentially lethal repercussions.

HSPB7, also referred to as cvHSP (18), has not previously been systematically examined for genetic variation. However, because an array-based analysis of approximately 2,000 candidate cardiovascular genes found a significant association between the intronic *HSPB7* SNP rs1739843 and heart failure (T.P. Cappola and G.W. Dorn II, unpublished observations), we included it in our resequencing study. We detected 39 *HSPB7* SNPs, of which 18 were either rare or observed exclusively in the African American subgroup (Table 2); to our knowledge, all were previously unreported. Of the *HSPB7* SNPs, 12 were associated with systolic heart failure, and these associations were independently validated in a second heart failure population. HSPs are upregulated in response to stress and exhibit a variety of cytoprotective effects through their function as molecular chaperones (40). A nonsynonymous polymorphism of HSP20 that lacks its characteristic cytoprotective effects has been described previously (41). However, none of the associated *HSPB7* SNPs changed an amino acid in HSPB7. To determine whether a functionally significant noncoding SNP might account for the association with heart failure, we examined the most recent list from the miRBase database of all the genomic locations of known human miR precursors; none were within 10 Mb of *HSPB7* (our unpublished observations). We also subjected the area we sequenced to TargetScan analysis, which showed only 1 SNP (that occurred only in African Americans and was not significantly associated in the case-control analysis) in a poorly conserved binding site for hsa-miR-1207-5p (our unpublished observations). Finally, we looked at proximity of the *HSPB7* SNPs to splice acceptor and splice donor sites; none were closer than 50 nucleotides (our unpublished observations). However, because the associated *HSPB7* SNPs were in linkage disequilibrium within a haplotype block that extends beyond the *HSPB7* gene, we speculate that the causative genetic event may be in another gene or in the intergenic region. Further examination of this locus by resequencing may help establish a mechanism for the genetic findings.

**Nonsynonymous signaling gene SNPs.** A major advantage of exomic sequencing is the potential to identify rare gene variants typically

missed by linkage analysis with microsatellite markers or marker SNPs. Any specific rare SNP, by definition, affects a small proportion of individuals and therefore tends to be discounted. As a class, however, rare SNPs can be collectively common, as evidenced by our observation that, of the 129 SNPs identified in this study, 74 (57.4%) had a prevalence of less than 1%. Because pathological effects of variant gene products produce selective pressure against sequence variation, a high proportion of rare nonsynonymous SNPs can be detrimental (27) and are therefore particularly important to catalog. Here, we identified 23 novel nonsynonymous SNPs (Supplemental Table 6) that — although they were not associated with increased risk for development of systolic heart failure in this study — may merit closer examination as potential modifiers of sporadic cardiomyopathy or of other diseases. To that end, using bioinformatics to classify nonsynonymous SNPs by their probability for pathological impact provides a mechanism for initial titration of SNPs for further mechanistic evaluation.

**Limitations.** Pooled sequencing, as used herein, can rapidly generate accurate SNP frequency data for a large study cohort, but does not provide genetic information at the individual level. Consequently, the data are not readily useful for assessment of relative risk. However, an estimate of carrier prevalence can be generated if one assumes that the SNPs conform to the predictions of Hardy-Weinberg equilibrium (42). With this caveat, the *HSPB7* SNPs identified herein increase the odds of developing systolic heart failure by an average of  $1.55 \pm 0.02$ -fold (Supplemental Table 7).

Pooled exomic sequencing is optimized for large-scale comparisons of allele frequencies between, or within, populations, as with case-control studies of disease or studies comparing populations of genetic/geographic ancestry. We have found that the technique and data analysis pipeline can be modified to incorporate individual PCR amplification and DNA barcoding prior to DNA pooling to obtain personal sequence, while taking advantage of the high throughput and efficiencies of pooled resequencing on the Illumina Genome Analyzer II (G.W. Dorn II and S.J. Matkovich, unpublished observations).

The nature of population genetics research, either using high-density genome-wide SNP microarrays to identify loci linked with disease or using targeted subgenome resequencing to identify specific SNPs associated with disease, is that the significant genetic events do not necessarily define the causal genetic events. Our study provides an example of this, in that multiple replicated and validated *HSPB7* SNPs associated with development of systolic heart failure in Caucasians do not, as far as we can determine using available bioinformatics resources, affect gene or protein function. Thus, the mechanism for these findings is unclear and will require both additional resequencing and case-control studies to identify linked SNPs that may be causal and wet laboratory work to define the functional impact of those SNPs at the cellular and organism levels.

Our SNP discovery efforts were performed in Caucasians and African Americans, and the heart failure case-control study was limited to systolic heart failure in Caucasians. The impact of our findings was not addressed in groups with different geographic ancestry, such as Hispanic and Asian populations, although the approach described herein would be ideal to do so.

In summary, we have demonstrated the accuracy and utility of pooled sequencing to identify previously unknown rare gene variants, to compare polymorphism expression between different ancestral populations, and to identify novel associations between *HSPB7* polymorphisms and sporadic systolic heart failure in Cau-



## research article

casians. This approach will be generally useful to identify disease-associated SNPs in gene-ontology clusters informed by array-based genome-wide association studies and by existing databases of gene expression or as suggested by biological rationale.

### Methods

**Study subjects.** Human study protocols were approved by the Institutional Review Boards of the University of Cincinnati or the University of Pennsylvania. All subjects provided written informed consent. Subjects with systolic heart failure were recruited, from patients presenting to the heart failure referral program at the University of Cincinnati or the University of Pennsylvania, into 1 of 2 longitudinal studies of heart failure genomics funded by the National Heart, Lung, and Blood Institute (NHLBI; P50 HL77101 and R01 HL88577). Enrollment criteria were age 18–80 years, clinical diagnosis of systolic heart failure, and documented abnormal left ventricular function (ejection fraction <40%) by noninvasive cardiac imaging. African American inclusion at greater than 25% of the total cohort was part of the NHLBI-approved study design. Nonaffected controls, in whom clinical heart failure was not present, echocardiographic left ventricular function was normal (ejection fraction >50%, no wall motion abnormalities), and there was no evidence for clinically significant coronary artery disease, were from the greater Cincinnati and Philadelphia metropolitan areas. DNA from a total of 2,606 individuals was studied in the primary cohort: 1,117 Caucasian heart failure subjects, 625 Caucasian nonaffected controls, 628 African American heart failure subjects, and 236 African American nonaffected controls. The secondary replication cohort consisted of 859 Caucasian systolic heart failure cases and 311 Caucasian nonaffected controls. Detailed clinical features of these heart failure study populations were recently reported (20).

**DNA preparation, pooling, and amplification.** Genomic DNA was isolated and extracted using the Gentra Puregene genomic DNA purification kit (Qiagen) and individually quantified by PicoGreen (Invitrogen) fluorescence in a 96-well format. For our initial validation experiments, 25 ng DNA from each individual was combined into pools, as shown in Supplemental Figure 1. DNA segments containing the regions of interest were individually PCR amplified using specific primers (Supplemental Table 8) and PfuUltra high-fidelity polymerase (Stratagene) from each of the pools, using an average of 30 diploid genomes (approximately 0.2 ng DNA) per individual as input into a total of 7 PCR reactions that included 9,389 bases from 4 genes: *ADRA1A*, *ADRB2*, *HSPB7*, and *PLN* (Supplemental Table 8). Preparation of the PCR products for sequencing on the Illumina Genome Analyzer I or II was as described previously (11). Briefly, amplicons were purified from primers and residual nucleotides by Qiaquick column separation (Qiagen), combined into mixtures containing an equivalent number ( $1 \times 10^{11}$ ) of molecules of each amplicon, and concatenated overnight at 22°C with T4 DNA ligase and T4 polynucleotide kinase (New England Biolabs) in the presence of 15% (w/v) polyethylene glycol, MW8000 (Sigma-Aldrich). pBluescript control DNA amplicons were included in ligations to monitor base calling accuracy. After 10-fold dilution in buffer PB (Qiagen), random fragmentation by sonication (Diagenode Bioruptor XL), and purification on Qiaquick columns (Qiagen), the DNA was end repaired, ligated to Illumina sequencing adapters, and prepared for analysis per Illumina protocol. Individual resequencing of *ADRA1A* exon 2 in the entire primary cohort was performed by dye terminator sequencing using an ABI 3730xl capillary sequencer (17, 43).

**Sequence analysis.** As previously demonstrated, the likelihood of sequencing errors across the 36-bp Illumina read across bases 1–12 averages 0.00065, rises dramatically after base 12, and can vary significantly from run to run (11). To model these errors, a 1,655-bp amplicon from the pBluescript backbone was incorporated into the ligation of PCR products and used to define an error model for each sequencing run. Sequencing output

was aligned against an annotated reference gene sequence, consisting of only the regions amplified, downloaded from the UCSC Genome Browser (44). Illumina reads were aligned against the reference by allowing 2 mismatches or fewer for each read. Any read with more than 2 mismatches or that aligned to multiple locations in the reference sequence was eliminated from SNP calling. SNPSeeker used only the first 12 highly accurate bases of each read along with a second-order dependency model to identify SNPs in all regions of interest. Therefore, whereas all 36 bases of each read were used for alignment to reference, only bases 3–12 of each 36-base read were effectively used for SNP identification. Because a single allele occurs at a frequency of 0.002 in a pool of 250 individuals (and at a frequency of 0.005 in a pool of 100 individuals), a singleton will be present at a frequency at least about 3-fold more prevalent than error in a given sequencing lane. We defined common SNPs as those occurring at a frequency of at least 0.01 in subjects of similar geographic ancestry, whereas the lower limit of detection for rare SNPs was a frequency of 0.0003 for all Caucasians (1 allele in the entire cohort of 1,742 individuals) and a frequency of 0.0006 for all African Americans (1 allele in 864 individuals).

For each gene, all available sequences were downloaded from dbSNP (19) and mapped onto an annotated reference gene sequence downloaded from the UCSC Genome Browser (44). Names were assigned according to the nomenclature suggested by the Human Genome Variation Society (45), in which the +1 position denotes the translation initiation codon, and numbering of SNPs within introns depends on proximity to the nearest exon.

**Comparative genomic analysis.** SIFT (28) and PolyPhen (29) analyses were conducted with the default parameter settings.

**Statistics.** Allele frequencies for all SNPs were compared with a Pearson's correlation coefficient. To assess differences in clinical variables, 2-tailed Student's *t* test was used. Fisher exact test or adjusted  $\chi^2$  test (see below) were used to compare allele frequencies. The *P* level threshold for significance in the primary heart failure case-control analysis was  $P < 0.0014$ , using a Bonferroni correction for multiple testing ( $n = 37$  common SNPs at  $\alpha = 0.05$ ). Prior to the replication study, linkage between the associated *HSPB7* SNPs was assessed. Allele frequency correlations across pools and eigenvalues were calculated using the R package per Gao et al. (46), and these were used to derive the effective number of independent SNPs ( $M_{\text{eff}}$ ) by the equation  $M_{\text{eff}} = 1 + [(M - 1) (1 - [\text{Var}(\lambda_{\text{obs}})]/M)]$ , in which *M* is the total number of variables in the matrix and  $\text{Var}(\lambda_{\text{obs}})$  is the variance of the observed eigenvalues (47). The calculated  $M_{\text{eff}}$  of 2.37 was then used in a Bonferroni correction to derive a threshold *P* value of  $P < 0.0212$  ( $\alpha = 0.05$ ). To account for variance due to DNA pooling, an estimate of pooling error was calculated according to Visscher and Le Hellard (48). The resulting error factor of 0.014 was then used to adjust the  $\chi^2$  test statistic.

### Acknowledgments

The present work was supported by Cardiac Translational Implementation Program (CTRIP) grant RC2 HL102222 from the NHLBI and Office of the Director (OD), NIH; by NIH grant P50 HL077101 to G.W. Dorn II; by NIH grant R01 HL088577 to T.P. Cappola; by National Center for Research Resources (NCRR), NIH, grant UL1 RR024992; and by Ruth L. Kirschstein National Research Service Award T32 HD043010 to T.E. Druley.

Received for publication July 8, 2009, and accepted in revised form October 14, 2009.

Address correspondence to: Gerald W. Dorn II, Washington University Center for Pharmacogenomics, 660 S. Euclid Ave., Campus Box 8220, St. Louis, Missouri 63110, USA. Phone: (314) 362-4892; Fax: (314) 362-8844; E-mail: [gdnorn@dom.wustl.edu](mailto:gdnorn@dom.wustl.edu).



1. Online Mendelian Inheritance in Man [database online]. Baltimore, MD: Johns Hopkins University Press; 2009. <http://www.ncbi.nlm.nih.gov/omim/>. Accessed November 5, 2009.
2. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
3. Chin L, Gray JW. Translating insights from the cancer genome into clinical practice. *Nature*. 2008;452(7187):553–563.
4. Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet*. 2008;17(R2):R116–R121.
5. McPherson R, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316(5830):1488–1491.
6. Zhu X, et al. Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet*. 2005;37(2):177–181.
7. Kurtz TW, Spence MA. Genetics of essential hypertension. *Am J Med*. 1993;94(1):77–84.
8. Florez JC, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet*. 2003;4:257–291.
9. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881–888.
10. The 1000 Genomes Project Data Release [database online]. 2009. <http://www.1000genomes.org/>. Accessed November 5, 2009.
11. Druley TE, et al. Accurate detection and quantification of rare allelic variants from the pooled genomic DNA of 1111 individuals. *Nature Methods*. 2009;6(4):263–265.
12. Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. *Nat Rev Genet*. 2004;5(5):335–344.
13. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53–59.
14. Lanfear DE, et al. Beta2-adrenergic receptor genotype and survival among patients receiving beta-blocker therapy after an acute coronary syndrome. *JAMA*. 2005;294(12):1526–1533.
15. Liggett SB, et al. The Ile164 beta2-adrenergic receptor polymorphism adversely affects the outcome of congestive heart failure. *J Clin Invest*. 1998;102(8):1534–1539.
16. Schmitt JP, et al. Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science*. 2003;299(5611):1410–1413.
17. Haghghi K, et al. A mutation in the human phospholamban gene, deleting arginine 14, results in lethal, hereditary cardiomyopathy. *Proc Natl Acad Sci U S A*. 2006;103(5):1388–1393.
18. Krief S, et al. Identification and characterization of cvHsp. A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues. *J Biol Chem*. 1999;274(51):36592–36600.
19. Single nucleotide polymorphism [database online]. 2008. <http://www.ncbi.nlm.nih.gov/projects/SNP/>. Accessed November 5, 2009.
20. Cresci S, et al. Clinical and genetic modifiers of long-term survival in heart failure. *J Am Coll Cardiol*. 2009;54(5):432–434.
21. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009;55(4):641–658.
22. O'Connell TD, et al.  $\alpha$ 1-adrenergic receptors prevent a maladaptive cardiac response to pressure overload. *J Clin Invest*. 2006;116(4):1005–1015.
23. Liggett SB. Polymorphisms of beta-adrenergic receptors in heart failure. *Am J Med*. 2004;117(7):525–527.
24. Kawase Y, Hajjar RJ. The cardiac sarcoplasmic/endoplasmic reticulum calcium ATPase: a potent target for cardiovascular diseases. *Nat Clin Pract Cardiovasc Med*. 2008;5(9):554–565.
25. Rathmore SS, et al. Race, quality of care, and outcomes of elderly patients hospitalized with heart failure. *JAMA*. 2003;289(19):2517–2524.
26. Yancy CW, Strong M. The natural history, epidemiology, and prognosis of heart failure in African Americans. *Congestive Heart Failure*. 2004;10(1):15–8.
27. Fay JC, Wyckoff GJ, Wu CI. Positive and negative selection on the human genome. *Genetics*. 2001;158(3):1227–1234.
28. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res*. 2002;12(3):436–446.
29. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30(17):3894–3900.
30. Freeman K, Nakao K, Leinwand LA. Low sequence variation in the gene encoding the human beta-myosin heavy chain. *Genomics*. 2001;76(1–3):73–80.
31. Tsai CT, Lai LP, Hwang JJ, Chiang FT. Molecular genetics of atrial fibrillation. *J Am Coll Cardiol*. 2008;52(4):241–250.
32. Hannenhalli S, et al. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation*. 2006;114(12):1269–1276.
33. Gu D, et al. Association of alpha1A adrenergic receptor gene variants on chromosome 8p21 with human stage 2 hypertension. *J Hypertens*. 2006;24(6):1049–1056.
34. Iacoviello M, et al. Alpha- and beta-adrenergic receptor polymorphisms in hypertensive and normotensive offspring. *J Cardiovasc Med (Hagerstown)*. 2006;7(5):316–321.
35. Pacanowski MA, et al. Adrenergic gene polymorphisms and cardiovascular risk in the NHLBI-sponsored Women's Ischemia Syndrome Evaluation. *J Transl Med*. 2008;6:11.
36. Brodde OE.  $\beta$ -1 and  $\beta$ -2 adrenoceptor polymorphisms: Functional importance, impact on cardiovascular diseases and drug responses. *Pharmacol Ther*. 2008;117(1):1–29.
37. Brittsan AG, Kranias EG. Phospholamban and cardiac contractile function. *J Mol Cell Cardiol*. 2000;32(12):2131–2139.
38. Haghghi K, et al. Human phospholamban null results in lethal dilated cardiomyopathy revealing a critical difference between mouse and human. *J Clin Invest*. 2003;111(6):869–876.
39. DeWitt MM, MacLeod HM, Soliven B, McNally EM. Phospholamban R14 deletion results in late-onset, mild, hereditary dilated cardiomyopathy. *J Am Coll Cardiol*. 2006;48(7):1396–1398.
40. Young JC, Agashe VR, Siegers K, Hartl FU. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol*. 2004;5(10):781–791.
41. Nicolaou P, et al. Human mutation in the anti-apoptotic heat shock protein 20 abrogates its cardioprotective effects. *J Biol Chem*. 2008;283(48):33465–33471.
42. Yang Y, et al. Efficiency of single-nucleotide polymorphism haplotype estimation from pooled DNA. *Proc Natl Acad Sci U S A*. 2003;100(12):7225–7230.
43. Liggett SB, et al. A GRK5 polymorphism that inhibits  $\beta$ -adrenergic receptor signaling is protective in heart failure. *Nat Med*. 2008;14(5):510–517.
44. Kuhn RM, et al. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res*. 2009;37(Database issue):D755–D761.
45. den Dunnen J. Nomenclature for the description of sequence variations. <http://www.hgvs.org/mutnomen>. Accessed November 5, 2009.
46. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008;32(4):361–369.
47. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004;74(4):765–769.
48. Visscher PM, Le HS. Simple method to analyze SNP-based association studies using DNA pools. *Genet Epidemiol*. 2003;24(4):291–296.