Genomic analysis reveals poor separation of human cardiomyopathies of ischemic and nonischemic etiologies

Ruprecht Kuner,1,6 Andreas S. Barth,2,3,6 Markus Ruschhaupt,1,5 Andreas Buness,3 Ludwig Zwermann,3 Eckart Kreuzer,4 Gerhard Steinbeck,3 Annemarie Poustka,1 Holger Sültmann,1,6 and Michael Nabauer3,6

1 Division of Molecular Genome Analysis, German Cancer Research Center, Heidelberg, Germany; 2 Division of Cardiology, Johns Hopkins University, Baltimore, Maryland; and 3 Department of Medicine I, and 4 Department of Cardiac Surgery, University Hospital Grosshadern, Ludwig Maximilians University and 5 Department of Medical Informatics, Biometrics, and Epidemiology, Ludwig Maximilians University, Munich, Germany

Submitted 20 December 2007; accepted in final form 19 April 2008

HEART FAILURE due to dilated cardiomyopathy, characterized by impaired contractile function of one or both ventricles and dilatation of the cardiac chambers, is a leading cause of hospitalization and death in Western countries (14). Most frequently, it develops in the setting of coronary atherosclerosis, with ischemic cardiomyopathy (ICM) accounting for >50% of cases of heart failure (2). The second largest heart failure entity is dilated cardiomyopathies of nonischemic origin (NICM), comprising various subgroups such as valvular and hypertensive heart disease, viral myocarditis, hereditary cases, toxic forms (e.g., alcohol and anthracycline cardiomyopathy), as well as a subset of idiopathic cases in which no specific cause can be determined.

In general, pharmacological treatment of overt heart failure is largely based on functional New York Heart Association (NYHA) status and differs little with respect to the underlying disease etiology. However, it has become increasingly clear that long-term prognosis and the response to certain pharmacological modalities may very well differ between ICM and NICM. For instance, response to digoxin and amiodarone was found to be better in nonischemic patients (1, 30). Likewise, large epidemiologic studies have shown that heart failure of ischemic origin is associated with a worse outcome compared with NICM (2, 18). Therefore, the differentiation between ischemic and nonischemic cases is clinically highly relevant but relies, so far, mainly on the results of a coronary angiography. While this invasive technique will correctly confirm (or rule out) significant coronary atherosclerosis in most cases, it might be misleading in up to 20% of patients who present after a myocardial infarction with spontaneous recanalization of the culprit coronary artery. Therefore, these patients will be misclassified as “nonischemic” in the absence of additional significant coronary stenoses (40). In this respect, genomic studies hold great promise to improve our current classification schemes. While we have recently reported a common and robust gene expression profile characteristic of failing human ventricular myocardium across >100 NICM and nonfailing (NF) samples of four independent microarray studies (4), it is currently less clear whether microarray-based signatures can also be used to infer the etiology of heart failure in end-stage NICM and ICM. Several pilot studies have suggested differences in genomic signatures between different etiologies of heart failure (7, 21, 22, 33), while others have failed to clearly differentiate between ischemic and idiopathic dilated cardiomyopathy on the basis of transcript profiling (12, 31). Therefore, we performed class prediction analysis independently in one cDNA and two publicly available high-density oligonucleotide microarray studies, comprising a total of 279 human myocardial samples. We then applied potential classifiers identified in a single study to the remaining data sets to test whether etiology-specific patterns of gene expression are discernible in different heart failure groups.

Genomic analysis reveals poor separation of human cardiomyopathies of ischemic and nonischemic etiologies. Physiol Genomics 34: 88–94, 2008. First published April 22, 2008; doi:10.1152/physiolgenomics.00299.2007.—Clinically, the differentiation between ischemic (ICM) and nonischemic (NICM) human cardiomyopathies is highly relevant, because ICM and NICM differ with respect to prognosis and certain aspects of pharmacological therapy, despite a common final phenotype characterized by ventricular dilatation and reduced contractility. So far, it is unclear whether microarray-based signatures can be used to infer the etiology of heart failure. Using three different classification algorithms, we independently analyzed one cDNA and two publicly available high-density oligonucleotide microarray studies comprising a total of 279 end-stage human heart failure samples. When classifiers identified in a single study were applied to the remaining studies, misclassification rates >25% for ICM and NICM specimens were noted, indicating poor separation of both etiologies. However, data mining of 458 classifier genes that were concordantly identified in at least two of the three data sets points to different biological processes in ICM vs. NICM. Consistent with the underlying ischemia, cytokine signaling pathways and immediate-early response genes were overrepresented in ICM samples, whereas NICM samples displayed a deregulation of cytoskeletal transcripts, genes encoding for the major histocompatibility complex, and antigen processing and presentation pathways, potentially pointing to immunologic processes in NICM. Overall, our results suggest that ICM and NICM exhibit substantial heterogeneity at the transcriptomic level. Prospective studies are required to test whether etiology-specific gene expression patterns are present at earlier disease stages or in subsets of both etiologies.

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
end-stage human heart disease of ischemic and nonischemic origin.

MATERIALS AND METHODS

cDNA microarray study. We performed a genomewide expression study that comprised septal myocardial samples from heart failure patients with end-stage NICM (n = 16) and ICM (n = 10). While the latter patients had a severe two- or three-vessel coronary disease (≥75% stenosis) and/or a history of myocardial infarction, coronary angiography showed no signs of significant (>50%) stenoses in NICM subjects. Detailed patient characteristics are given in Supplemental Table S1.1 All patients gave written informed consent. The investigation was approved by the Institutional Review Board and is in accordance with the Helsinki Declaration of 1975, as revised in 1983.

As a two-color microarray experiment, each NICM and ICM sample was labeled with Cy3 and hybridized against a common pool of RNA from 15 NF hearts labeled with Cy5. RNA amplification, labeling, and hybridization to RZPD Unigene 3.1 cDNA microarrays (37.5 K) were carried out as described previously (4). Preprocessing and most of the statistical analysis were done with Bioconductor (www.bioconductor.org). The cDNA microarray raw data were normalized with the “variance stabilization” method (VSN; Ref. 20) and arrayMagic software package (10). Hierarchical clustering of genes and samples was based on the Euclidean distance measure. The platform iCHIP (Integration Centre of High Throughput Experiments: www.ichip.de) was used to export the microarray data via MAGE-ML to the repository ArrayExpress of the European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress, accession number E-CVDE-1).

Etiologic classification of three microarray studies of human heart failure. We classified NICM and ICM samples in one cDNA (E-CVDE-1) and two publicly available high-density oligonucleotide microarray studies (NCBI Gene Expression Omnibus database accession numbers GSE5406 and GSE1145). Data set GSE5406 represents the largest microarray study of human heart failure so far, comparing 108 ICM and 86 NICM samples hybridized to Affymetrix HG-U133A arrays (19). Data set GSE1145 consists of 32 ICM and 27 NICM samples hybridized to Affymetrix HG-U133 2.0 plus arrays (www.cardiogenomics.org). In summary, the present meta-analysis comprised a total of 150 ICM and 129 NICM samples from three independent microarray studies. More detailed information regarding these studies is provided in Table 1. Affymetrix probes and cDNA clones were merged between different microarray platforms with Bioconductor annotation packages (www.bioconductor.org).

Three different classification methods [prediction analysis for microarrays (PAM; Ref. 34), random forest (RF; Ref. 9), and penalized logistic regression (PLR; Ref. 41)] were used to classify ICM and NICM samples in three independent microarray data sets. PAM uses denoised versions of the centroids as prototypes for each class. In the RF method, different decision trees are grown by introducing a random element into their construction. First, each tree makes a prediction of one class for each observation, and then these predictions are combined to get the final classification result. PLR combines a logistic regression model with a ridge penalty to classify microarray data. To estimate the fraction of misclassified samples, a complete cross-validation implemented in the Bioconductor package “MCRes- timate” (28) was applied. This complete cross-validation includes an additional gene filtering step based on the genes with the highest variance. After estimating the misclassification rate in all three studies by the three aforementioned methods of classification, we then applied the most successful classifier (GSE1145, RF) as well as a published classifier of ICM vs. NICM (22) to the remaining microarray studies.

Table 1. Characterization of human microarray studies focusing on different etiologies of heart failure

<table>
<thead>
<tr>
<th>Accession No. (public repository)</th>
<th>Present study</th>
<th>Raw Data Accessible</th>
<th>Technical Replicates</th>
<th>Normalization</th>
<th>Filtering Criterion</th>
<th>Autopsy Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-CVDE-1 (Array Express)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>VSN</td>
<td>Absent/present calls</td>
<td>No</td>
</tr>
<tr>
<td>GSE1145 (GEO)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>RMA</td>
<td>Absent/present calls</td>
<td>No</td>
</tr>
<tr>
<td>GSE5406 (GEO)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>RMA</td>
<td>Absent/present calls</td>
<td>No</td>
</tr>
<tr>
<td>GSE1869 (GEO)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>RMA</td>
<td>Absent/present calls</td>
<td>No</td>
</tr>
</tbody>
</table>

ICM, ischemic cardiomyopathy; NICM, nonischemic cardiomyopathy; VSN, variance stabilization; RMA, robust multiarray average; GEO, Gene Expression Omnibus.

1 The online version of this article contains supplemental material.
Data mining was carried out with the classifiers derived from the three aforementioned studies and with 458 classifier genes that were concordantly identified in at least two of the three data sets. Statistical analysis of these genes was based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and the Gene Ontology (GO) domains “cellular component,” “biological process,” and “molecular function,” implemented in “Fatigo” (3). Ingenuity Pathways Knowledge Base (Ingenuity, Mountain View, CA), the largest manually curated gene database, was used to identify literature-based gene-gene interaction networks. The utility of this database for microarray data of myocardial samples has been demonstrated previously (38).

RESULTS

Classification of nonischemic and ischemic cardiomyopathies. Applying three different classification algorithms to our own cDNA microarray data set (E-CVDE-1) and two larger, publicly available microarray data sets (GSE1145 and GSE5406), we found that on average one-fourth of the samples were misclassified (range 14–32%; Fig. 1 and Supplemental Figs. S1, S2, and S3). Because the definition of incorrectly classified samples was based on 50 resamplings, Supplemental Figs. S1, S2, and S3 show a graphical representation of the results of the cross-validation for each individual sample (i.e., in Supplemental Fig. S1, samples 93, 101, and 103–106 were consistently misclassified, while sample 102 shows borderline classification results, with an equal number of correct and incorrect classifications). Furthermore, these graphs demonstrate that the results of the classification were not very stable even in those cases where an apparent low misclassification rate was computed, e.g., with RF (14%) and PLR (17%) in data set GSE1145. This is best exemplified for the PLR method in GSE1145 (Supplemental Fig. S2), where one can appreciate a high variation even for those samples that were correctly classified. The variability suggests that the classifier itself is not very robust. This is supported by the fact that the results of the best classifier (correct classification of 86% of samples with RF in data set GSE1145) could not be reproduced in independent data sets, because misclassification rates were 34% and 35% in data sets GSE5406 and E-CVDE-1, respectively. As a limitation of the analysis of different microarray platforms, the 509 nonredundant classifier genes could only be matched to a subset of identifiers in the two other data sets (GSE5406: 414 genes and E-CVDE-1: 289 genes). To exclude that the poor classification result in data sets GSE5406 and E-CVDE-1 might be due to the lack of genes important for the classification process in these data sets, we applied the RF algorithm in GSE1145, using 414 genes available in GSE5406 and 289 genes available in data set E-CVDE-1. Because of overfitting, these results cannot be compared with the initial classification result for the data set. Therefore, we also performed the RF classification based on the 509 nonredundant classifier genes. There was no difference between the three resulting misclassification rates (data not shown). This suggests that the poor classification results obtained with the RF classifier from GSE1145 in data sets GSE5406 and E-CVDE-1 is not due to the lack of genes important for the classification algorithm. To further support this notion, we performed an additional analysis in the two largest data sets, GSE1145 and GSE5406, encompassing >90% of all 279 samples included in the present article. Based on ENTREZ IDs, we first determined the genes common to both data sets. Second, we applied the RF classification procedure in GSE1145, using only probe sets that were shared between both data sets. Supplemental Figure S5 shows that the classification results obtained in data set GSE1145 (misclassification rate 14%) are similar to our previous analysis using the complete data set (Fig. 1). However, application of this classifier to GSE5406 still yielded a poor classification, with a misclassification rate of 34%.

Likewise, we found no clear evidence that sex influences the misclassification rates for NICM or ICM samples (8). Even though men are in general overrepresented in both etiologies (across all studies ~60% and ~85% of subjects were male in the NICM and ICM groups, respectively), sex did not play a major role in the results of the current classification for NICM or ICM samples (Supplemental Table S1). This was also supported by a separate analysis of myocardial samples of male patients only that yielded similar results compared with analysis of the entire data set (Supplemental Fig. S6).

In an additional classification analysis, we examined whether a published 90-gene classifier for ICM and NICM (22) can correctly distinguish ICM from NICM in the three aforementioned microarray data sets. Here, the classifier genes were only reduced in data set E-CVDE-1 (35 genes). However, 40%
of 279 myocardial samples were misclassified again, indicating that results from this study cannot be extrapolated to independent cardiomyopathy data sets (Fig. 1).

In contrast, by applying our previously described genomic classifier for heart failure (4) to the largest data set (GSE5406, Supplemental Fig. S4), we now achieved successful classification between NF and failing myocardium in a total of 347 of 360 samples, yielding an overall prediction accuracy of 96.4%, which is in sharp contrast to the high misclassification rate between NICM and ICM in the same microarray data set.

Distinct biological processes in ischemic and nonischemic dilated cardiomyopathies. To obtain etiology-specific transcriptional patterns in NICM vs. ICM, we performed GO and KEGG pathway analyses with classifiers derived from each data set separately. Furthermore, to define a common, minimal denominator across independent data sets, we also performed a GO analysis with 458 classifier genes that were concordantly deregulated in at least two of the three microarray studies (Fig. 2A and Supplemental Tables S2 and S3). As a result, we found an overrepresentation of cytoskeletal transcripts, genes encoding for the major histocompatibility complex, and antigen processing and presentation pathways in NICM samples. In contrast, the GO groups of “inflammatory response,” “cytokine activity,” and “G protein-coupled receptor binding” were enriched in ICM samples (Fig. 2B). Further analysis of the classifier genes characteristic of ICM samples revealed an overrepresentation of these transcripts in a prominent gene-gene interaction network associated with the immediate-early response genes FOS and JUNB (Fig. 3A). Of note, many of the genes comprised in this gene network have previously been implicated in human cardiomyopathies, including CCL2 (4), ATF3 (4), SFRP4 (4), LAMP2 (39), IGFBP (6), and TIMP1 (5). Upregulation of these ICM classifier genes could be linked to a subset of ICM specimens, as suggested by the hierarchical clustering in data set GSE1145 (Fig. 3B) with separation of two distinct clusters comprising predominantly ICM samples on the left (11 ICM vs. 3 NICM) and a cluster with a nearly equal number of ICM and NICM samples on the right.

Notably, t-test-based methods did not allow a comprehensive analysis of the different etiologies since statistical analysis of microarrays (SAM) analysis (35) resulted only in a list of 13 genes in the largest data set (GSE5406) with negligible overlap to the other microarray data sets (data not shown).

DISCUSSION
Classifying end-stage ischemic and nonischemic cardiomyopathies. Clinically, the differentiation into NICM and ICM is highly relevant because, in nonischemic cases, secondary causes (e.g., myocarditis and excessive alcohol consumption)
need to be identified, while ischemic patients, on the contrary, may benefit from revascularization and secondary preventive pharmacotherapy with lipid-lowering drugs and antiplatelet agents. However, correct classification between ischemic and nonischemic cases is difficult, as highlighted by a study using gadolinium-enhanced cardiovascular magnetic resonance imaging that suggested an incorrect assignment of ischemic and nonischemic heart failure etiology in 13% of heart failure patients (24). Likewise, the Assessment of Treatment with Lisinopril and Survival study (ATLAS) showed an incorrect classification of NICM and ICM cases in up to 28% based on autopsy data (36). In this respect, incorrect assignment of disease etiology based on coronary angiography would not only be misleading in patients in whom spontaneous revascularization of the coronary artery occurs after a myocardial infarction (40), but also in patients who are classified as NICM and who show major coronary findings at autopsy (36).

By convention, new diagnostic tests should be validated against an established gold standard. Because several clinical modalities including coronary angiography and magnetic resonance imaging (MRI) have been shown to lack the sensitivity to correctly classify the etiology of heart failure in all cases, the most appropriate reference standard is postmortem autopsy. As a limitation, autopsy data were only available for a subset of cases in our own cDNA microarray study (Supplemental Table S1) and have not been provided in any of the published studies dealing with different etiologies of heart failure (19, 22). Thus, in the present study, phenotypic classification was largely based on information routinely used by clinicians to assign patients to ischemic or nonischemic disease etiologies. However, it is of interest to note that even in those cases that could be diagnosed unequivocally as ICM, misclassification based on microarray data was common. As an example, 5 of 10 ICM samples were misclassified in the cDNA microarray study—despite the fact that these patients had a history of prior myocardial infarctions, coronary artery bypass graft surgeries, and three-vessel coronary disease on coronary angiograms. Several explanations are possible. First, a severe ICM is expected to exhibit significant heterogeneity at the molecular level, reflecting areas of nonischemic, ischemic, stunned, or hibernating myocardium, all of which show distinct genomic profiles in animal models and human heart failure (15–17). In agreement with this hypothesis, we noted significant heterogeneity of ischemia marker genes in samples from ICM patients (Fig. 3). The fact that those ischemia marker genes were also upregulated in some NICM samples goes along with the fact that coronary microvascular dysfunction has been shown to be prevalent in NICM as well (26), leading to defined areas of

**Fig. 3.** A: Ingenuity pathway analysis suggested an overrepresentation of ICM classifier genes in a gene-gene interaction network with the 2 immediate-early response genes FOS and JUNB. The genes shown in gray are derived from the ICM classifier list (Supplemental Table S3). The other genes were suggested from the literature as potential interaction partners. B: unsupervised clustering using Euclidean distance for the genes of the Ingenuity network in data set GSE1145. Red indicates upregulation and blue indicates downregulation of these genes. NICM samples are represented by yellow and ICM samples by green bars. Upregulation of these ICM classifier genes could be linked to a subset of ICM specimens, as the clustering suggested separation of samples into 2 distinct clusters, 1 on the left comprising predominantly ICM samples (11 ICM and 3 NICM) and 1 on the right with a nearly equal number of ICM and NICM samples.
Atherosclerosis is currently considered to be an inflammatory disease in which the initiation and progression of an atherosclerotic toward an unstable plaque is driven by leukocyte recruitment mediated by various inflammatory mediators, including chemotactic cytokines or chemokines (11). Consistent with this hypothesis, ICM samples were characterized by the functional GO classes of “inflammatory response,” “chemotaxis,” “cytokine activity,” and “myeloid cell differentiation” (Fig. 2). Ingenuity pathway analysis suggested a gene-gene interaction network centered around the immediate-early response genes FOS and JUNB. It has been shown that brief episodes of myocardial ischemia activate a distinct genetic program promoting cardioprotection and cell survival, with the immediate-early response genes being among the most consistent findings of ischemic preconditioning (15, 27).

In NICM, GO analysis suggested involvement of immune processes with upregulation of major histocompatibility complex genes. In line with this finding, KEGG pathway analysis revealed “antigen processing and presenting pathways” to be more prominent in NICM than ICM. Of note, idiopathic nonischemic dilated cardiomyopathy has been proposed as a disease with autoimmune features (25). The finding that immune processes were involved in both ICM and NICM is consistent with our previous study (4), in which we found a deregulation of the immune system to be characteristic of end-stage human failing myocardium. In addition to deregulation of immune processes in NICM, it is of particular interest to find deregulation of cytoskeletal and sarcomere transcripts as well (Supplemental Tables S2 and S3), especially as a wide variety of studies point to a link between mutations in cytoskeletal genes and the initiation of dilated cardiomyopathy (13). However, one has to keep in mind that our classification results and the low number of differentially expressed genes suggest substantial similarities in the transcriptomic response of ICM and NICM. Transcriptional differences between the two etiologies do exist, but these differences seem to be rather small when the analysis is based on the preassigned clinical categories. Thus prospective studies are required to identify specific subgroups of patients and test whether distinct transcriptomic patterns and related biological processes carry a different prognosis and require individualized therapies (i.e., platelet inhibition for NICM patients with an ischemic gene expression profile).

ACKNOWLEDGMENTS

The authors are thankful to Jürgen Eils and Chris Lawerenz from the Division of Theoretical Bioinformatics in the German Cancer Research Centre (DKFZ) regarding their support for iChip (Integration Centre of High Throughput Experiments; www.ichip.de).

GRANTS

This study was supported by a grant of the German Federal Ministry for Education and Research supporting the German National Genome Research Network [BMBF Grant 01GS0109 (G. Steinbeck, M. Nabauer); NGFN Grant 01GR0101 (R. Kuner, A. Buneß, H. Sültmann); NGFN Grant 01GR0459 (M. Ruschhaupt)] and a grant from the Deutsche Forschungsgemeinschaft [DFG; Grant BA 3341/1-1 (A. S. Barth)].

REFERENCES


