Transcriptional networks characterize ventricular dysfunction after myocardial infarction: A proof-of-concept investigation

Francisco Azuaje a,*, Yvan Devaux a, Melanie Vausort a, Céline Yvorra a, Daniel R. Wagner a,b

a Laboratory of Cardiovascular Research, Centre de Recherche Public – Santé, Luxembourg
b Division of Cardiology, Centre Hospitalier, Luxembourg

ARTICLE INFO

Article history:
Received 5 February 2010
Available online 23 May 2010

Keywords:
Myocardial infarction
Ventricular dysfunction
Angiogenesis
Translational bioinformatics
Systems medicine
Medical decision-support systems
Cardiovascular diseases

ABSTRACT

There is currently no method powerful enough to identify patients at risk of developing ventricular dysfunction after myocardial infarction (MI). We aimed to identify major mechanisms related to ventricular dysfunction to predict outcome after MI. Based on the combination of domain knowledge, protein–protein interaction networks and gene expression data, a set of potential biomarkers of ventricular dysfunction after MI was identified. Here we propose a new strategy for the prediction of ventricular dysfunction after MI based on “network activity indices” (NAI), which encode gene network-based signatures and distinguishes between prognostic classes. These models outperformed prognostic models based on standard differential expression analysis. NAI-based models reported high classification accuracy, with a maximum area under the receiver operating characteristic curve (AUC) of 0.75. Furthermore, the classification capacity of these models was validated by performing evaluations on an independent patient cohort (maximum AUC = 0.75). These results suggest that transcriptional network-based biosignatures can offer both powerful and biologically-meaningful prediction models of ventricular dysfunction after MI. This research reports a new integrative strategy for identifying transcriptional responses that characterize cardiac repair and for predicting clinical outcome after MI. It can be adapted to other clinical domains, such as those constrained by small molecular datasets and limited translational knowledge. Furthermore, it may reflect clinically-meaningful synergistic effects that cannot be identified by standard analyses.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Despite timely reperfusion therapy with balloon angioplasty or thrombolysis, a significant proportion of patients with acute myocardial infarction (MI) still develop left ventricular dysfunction and heart failure. This is associated with adverse clinical outcome [1]. Reperfusion of myocardial tissue is abnormal in these patients despite successful restoration of epicardial coronary blood flow. Dysfunction of the microvascular bed is the major mechanism behind abnormal tissue reperfusion in the setting of MI, and it is believed that angiogenic factors may restore microcirculation after reperfusion [2]. Indeed, several angiogenic factors have been measured in patients with acute coronary syndromes and levels of VEGF were found to predict prognosis [3]. In addition, it was observed that cardiac repair after cell transplantation is dependent on the presence of VEGF [4]. However, these observations and traditional biomarker discovery methods are based on standard statistical models and lists of biomarkers assumed to be independent of one another. Despite their relevance, they still focus on the behavior of individual markers based on relative expression levels across samples. This is insufficient to provide deeper clues into the mechanisms implicated in left ventricular dysfunction and angiogenesis after MI. This motivates novel approaches to understand the dynamics of angiogenesis and the prediction of ventricular dysfunction in post-MI patients.

The notion that a better understanding of clinical outcome after MI requires the analysis of biological networks is accepted. However, these observations and traditional biomarker discovery methods are based on standard statistical models and lists of biomarkers assumed to be independent of one another. Despite their relevance, they still focus on the behavior of individual markers based on relative expression levels across samples. This is insufficient to provide deeper clues into the mechanisms implicated in left ventricular dysfunction and angiogenesis after MI. This motivates novel approaches to understand the dynamics of angiogenesis and the prediction of ventricular dysfunction in post-MI patients.

There is currently no method powerful enough to identify patients at risk of developing ventricular dysfunction after myocardial infarction (MI). We aimed to identify major mechanisms related to ventricular dysfunction to predict outcome after MI. Based on the combination of domain knowledge, protein–protein interaction networks and gene expression data, a set of potential biomarkers of ventricular dysfunction after MI was identified. Here we propose a new strategy for the prediction of ventricular dysfunction after MI based on “network activity indices” (NAI), which encode gene network-based signatures and distinguishes between prognostic classes. These models outperformed prognostic models based on standard differential expression analysis. NAI-based models reported high classification accuracy, with a maximum area under the receiver operating characteristic curve (AUC) of 0.75. Furthermore, the classification capacity of these models was validated by performing evaluations on an independent patient cohort (maximum AUC = 0.75). These results suggest that transcriptional network-based biosignatures can offer both powerful and biologically-meaningful prediction models of ventricular dysfunction after MI. This research reports a new integrative strategy for identifying transcriptional responses that characterize cardiac repair and for predicting clinical outcome after MI. It can be adapted to other clinical domains, such as those constrained by small molecular datasets and limited translational knowledge. Furthermore, it may reflect clinically-meaningful synergistic effects that cannot be identified by standard analyses.

© 2010 Elsevier Inc. All rights reserved.
Network-based approaches to predicting gene–disease relationships have been explored in other areas. Recent studies have mainly focused on the prediction of potential drug targets and biomarkers in cancers. They have proposed algorithms for extracting sub-networks implicated in disease [5–7], and the detailed re-construction of regulatory networks to identify key signaling pathways [8]. These approaches in principle rely on relatively large collections of gene expression data or extensive sources of annotated interactions. In this paper, we assemble resources of PPI, gene expression and transcriptional associations implicated in processes relevant to angiogenesis in post-MI patients. Moreover, unlike previous research in post-MI ventricular dysfunction, we explore the potential of network-based prediction models for clinical decision-making support. We developed new prognostic models based on patient-specific expression synchronization of a small number of genes. These models capture network-based prognostic signatures, which are defined as “activity indices” and provide the basis for a novel prognostic strategy.

2. Methods

2.1. Study design and analysis framework

Our investigation consisted of four phases: data acquisition, integrative data mining, knowledge discovery and evaluation, and independent validation (Fig. 1). It has been suggested that angiogenesis may play a cardioprotective role in the setting of myocardial ischemia [4,9]. This motivated the representation of a network of PPI relevant to angiogenesis and MI in humans (step 1, Fig. 1). Also in the first phase, expression profiles of blood cells harvested at the time of mechanical reperfusion in two groups of 16 patients each (Table S1 in Supplementary Information) was determined using whole-genome oligonucleotide microarrays (step 2). Peripheral blood gene expression profiles have previously been used to detect allograft rejection in patients with heart transplantation and found to correlate with findings of endomyocardial biopsies [10]. Ejection fraction (EF) measured with echocardiography using Simpson’s method, 1 month after MI, was used to classify the patients into low EF (≤40%; median: 35, range: 20–40) or high EF (>40%; median: 63, range: 45–73) groups. Both groups of patients were age- and sex-matched and did not differ with respect to reperfusion time, final coronary flow, multivesSEL disease, history of previous infarction, cardiovascular risk factors or treatment. As expected, enzyme release was higher in the group of patients with impaired EF, which is consistent with impaired tissue perfusion. Of note, 6 patients in the low EF group died (37%) and 3 developed heart failure during a 2-year follow-up, while no patient died in the high EF group and only 1 developed heart failure. These data confirm the poor prognosis associated with ventricular dysfunction after MI.

In the second phase, clustering of the PPI network identified a set of genes significantly implicated in the regulation of growth and angiogenesis (step 3). From this set of genes, those found expressed in low and high EF patients by microarrays were used to construct co-expression networks specific to the EF groups (step 4). In the third phase, analysis of co-expression networks revealed potentially novel prognostic models (step 5), which were tested for their ability to classify patients into low or high EF groups (step 6). Our models exploited the application of “network activity indices” (NAIs) for patient classification. Finally, using an independent dataset obtained with multiplex PCR (step 7), classification models were validated (step 8).

2.2. Patients and microarrays

All patients had successful mechanical reperfusion and stenting of the infarct artery within 12 h of chest pain onset. All patients received Aspirin, Clopidogrel, Heparin and Abciximab. Two groups consisting each of 16 patients exhibiting different clinical outcome after MI were analyzed (Table S1 in Supplementary Information). One group of patients had a preserved left-ventricular (LV) systolic function and high EF 1-month after MI (>40%, median: 63, range: 45–73), and the other group showed impaired LV function and low EF (≤40%, median: 35%, range: 20–40). Total RNA was extracted from whole-blood cells by the PAXgene™ technology. A common reference RNA (Universal Human Reference RNA, Stratagene Europe, Amsterdam, The Netherlands) was used in conjunction with patient’s RNA. Messenger RNAs were amplified using the Amino Allyl MessageAmp™ kit (Ambion®, Cambridgeshire, United King-
dom). Five microgram of each amino allyl aRNA were labeled with Cy3 or Cy5 (Amersham, Buckinghamshire, United Kingdom). 750 ng of each labeled amino allyl aRNA (reference RNA or donor RNA) were combined and hybridized on four whole-genome oligonucleotide microarrays [11]. Spot finding and raw data quantification of microarrays were performed using the MAIA® freeware [12]. A Lowess non-linear normalization step was performed with the Acuity® software (Molecular Devices, Berks, UK). A filtering step was then performed to remove genes that were not present in at least three microarrays out of four. Filtering analysis was performed using the Significance Analysis of Microarrays (SAM) software which correlates gene expression with an external variable, such as the EF group [13]. Missing-value imputation was performed via a K-Nearest Neighbor algorithm normalization using 10 neighbours. Gene expression data are available at the GEO database under the accession number GSE11947 [14].

2.3. Independent cohort and multiplex PCR data

The independent cohort consisted of 62 patients with MI (Table S3 in Supplementary Information). From these, two groups were retrieved: one group with low EF 1-month after MI (<40%, median: 38%; range: 17–40) and one group with preserved EF (>40%, median: 55%; range: 42–86). One microgram of total RNA extracted from whole-blood cells by the PAXgene™ technology was reverse-transcribed using the Superscript® II reverse transcriptase (Invitrogen, Merelbeke, Belgium), Compatible five-plex PCR primers and dual-la

beled probes designed with the PrimerPlex software (Premier Bio
sot, Palo Alto, USA) were purchased from TIB MolBiol (Berlin,
Germany) [Supplementary Information]. Using the GeNorm algo
rithm (PrimerDesign, Southampton, UK), SF3a1 was determined as
housekeeping gene for normalization. The fluorophores for probe
labeling were selected to prevent emission spectra overlaps. PCR
was performed in a total volume of 20

| l containing: 1X PCR buffer II (Applied Biosystems, Halle, Belgium), 5 mM MgCl2, 2.5 U AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 300 μM dNTPs (Invitrogen), 4 μl of 1/10 dilutions of cDNA and primers/probes [Supplementary Information]. Multiplex quantitative real-time PCR was performed using the IQ5™ thermocycler (Biorad, Nazareth, Belgium). After initial denaturation for 7.5 min at 55°C, 50 cycles of 20 s at 95°C and 1 min at 60°C were performed. Each run included negative controls and a common control cDNA to check for inter-run variability. Expression levels were calculated by the ΔCt method using the housekeeping gene as reference.

2.4. Protein–protein interaction network construction and clustering analysis

The Entrez-Gene database [15] was used to retrieve a set of core genes known to be associated with angiogenesis in MI, using the keywords “human, heart, angiogenesis, myocardial, infarction”. Annotated PPI associated with these genes were retrieved from the Human Protein Reference Database together with other interacting proteins not included in the set of core genes [16]. The network was visualized and analyzed with Cytoscape [17]. A network clustering analysis was implemented to identify potential functional network modules. Clusters were identified by the (Cytoscape plug-in) MCODE network clustering algorithm [18]. Analyses of the significance of functional roles over-represented in these clusters were implemented with Fatigo under the Babelomics platform [19].

2.5. Construction and analyses of gene co-expression networks

Gene expression networks were constructed and analyzed with BioLayout Express3d [20]. Gene co-expression was calculated with the Pearson correlation coefficient. Results were unaffected when using the Spearman Rank correlation. The integration of gene expression data and co-expression networks was implemented with Cytoscape [17].

2.6. Statistical analysis and software tools

Comparisons between groups were performed with the SigmaPlot (v11.0) and the Statistica (v8.0) software. Statistical hypotheses testing on continuous numerical data were implemented with two-tailed t-tests. Proportions were compared using two-tailed Fisher’s exact tests. P values <0.05 were considered as statistically detectable. P values for the associations between network clusters and GO terms were estimated using (two-tailed) Fisher’s exact tests and corrected for account for multiple-hypotheses testing using the Benjamini & Hochberg adjustment procedure [19]. Classifiers were built and evaluated with the Weka system [21].

3. Results

3.1. Assembling and clustering analysis of a PPI network

Based on the assumption that angiogenesis may benefit tissue perfusion after MI, a PPI network of angiogenesis in human MI was established. The network was constructed by first extracting from the Entrez-Gene database [15] a set of 35 genes known to be relevant to angiogenesis and human MI (Material and Methods, and Data S1). PPIs associated with these 35 genes were obtained from a database of curated PPIs (Methods). The resulting network consisted of 556 nodes (proteins) and 686 edges (interactions) (Fig. S1 in Supplementary Information).

Clustering analysis of this network using different clustering parameters consistently identified a sub-network with 53 proteins highly specialized in growth regulation processes, as defined in the Gene Ontology (GO) [22]. This motivated us to define it as the “growth and regulation” module (Data S2). Of the initial set of 35 angiogenesis proteins, only 3 proteins (CTGF, TGFβ1 and VEGFA) were included in the cluster of 53. Among the genes encoding these 53 proteins, 33 genes were expressed in the blood samples as shown by our microarray analysis, and among these 33 genes, VEGFB and THBS1 were found differentially expressed between low and high EF patients as detected by SAM. In the PPI network, VEGFB only interacts with VEGFA, while THBS1 interacts with TGFβ1 and MMP9. Subsequent analyses focused on this set of 33 genes.

3.2. Construction of co-expression networks

Co-expression networks specific to each EF class were assembled. High EF and low EF networks were created using samples belonging to the corresponding class only. Nodes and edges represent genes and co-expression, respectively, above a specified correlation threshold (cT). Class-specific networks were constructed for different cT values. The number of nodes, edges and average connectivity kept constant for cT < 0.75 (Fig. S2 in Supplementary Information). Because of this and our interest in strong relationships, subsequent analyses concentrated on networks with cT ≥ 0.75. Their small sizes, i.e. the low number of genes and interactions they contain, also make them suitable for prognostic interpretation.

Fig. 2 depicts the high EF and low EF networks obtained with cT = 0.75, 0.80, 0.85. Differential expression synchronization patterns were found suggesting that clinical outcome after MI is linked to a specific coordination of gene expression. Relative tight
co-expression between VEGFB and DAXX on the one hand, and TNXB and LTBP4 on the other hand, exist in the high EF category. A correlation reduction was observed in the low EF group. Among the four potential biomarkers, only VEGFB was significantly differentially expressed ($t = 3.35, P = 0.004$). The other biomarkers did not show differential expression between low and high EF patients ($P > 0.05$). In high EF samples, the correlation between VEGFB and DAXX was equal to 0.78 (partial correlation controlling for the other gene pair, $P < 0.05, 95\%$ confidence interval (CI): [0.46, 0.93]). In low EF samples, their partial correlation was reduced to 0.29 ($P > 0.05, CI: [-0.69, -0.25]$). With regard to TNXB and LTBP4: in high EF samples, their partial correlation was 0.81 ($P < 0.05, CI: [0.53, 0.94]$), and in low EF was 0.71 ($P < 0.05, CI: [0.33, 0.90]$). These four potential biomarkers do not interact between them in the global PPI network. Each one has one interacting partner only: VEGFB with VEGFA, DAXX with TGFB1, TNXB with VEGFA, and LTBP4 with TGFB1. These observations motivated us to further explore the use of the pair-wise expression values of these genes to distinguish between EF classes.

3.3. Functional analysis of co-expression networks

A closer look at the composition of the co-expression networks showed functional differentiation. Fig. 3 shows examples of biological processes consistently over-represented in each EF class, as defined in the GO. All terms shown reported statistically detectable enrichments with $P < 0.05$, after correcting for multiple-testing. Statistical significance of the enrichments were estimated with (two-tailed) Fisher’s exact tests and corrected with the Benjamini & Hochberg adjustment procedure [19].

**Fig. 2.** Clinical outcome after MI is linked to a specific coordination of gene expression. Low and high EF patients are differentiated on the basis of gene expression synchronization. In these networks nodes and edges represent genes and their co-expression, respectively. An edge is established between two genes when their co-expression values are greater than the specified correlation threshold (cT). Co-expression networks for different cT show that expression of selected genes is differently coordinated between low and high EF patients. In patients with high EF, expression of TNXB and LTBP4, as well as DAXX and VEGFB, was synchronized at the time of reperfusion. In patients with low EF, this synchrony was lost.

**Fig. 3.** GO biological processes in which strong co-expression exist between the genes from Fig. 2. Grey squares indicate enriched GO category. All terms shown reported statistically detectable enrichments with $P < 0.05$, after correcting for multiple-testing. Statistical significance of the enrichments were estimated with (two-tailed) Fisher’s exact tests and corrected with the Benjamini & Hochberg adjustment procedure [19].
3.4. Prognostic classification based on co-expression networks

A crucial question is whether the EF class can be predicted in a patient-specific context. A possible approach is the matching between a sample and the obtained networks. To address this challenge, we mapped the 32 samples independently on each of the relevant networks shown in Fig. 2. In Fig. 5 an expression level in a sample is reflected by using a color-coded scale from green (low) to red (high expression) relative to the genes shown in the network. This and the network connectivity offered a method to compare associations in a sample in relation to an EF-specific network. Thus, for a mapped sample, linked nodes showing similar colors indicate good coordination between the genes in the mapped sample. On the other hand, linked nodes showing dissimilar colors indicate poor gene co-expression in the mapped sample (Fig. 5).

When samples from the high EF group are mapped onto high EF networks, between-node correlation is also reflected on the basis of the color-coded scheme (e.g. upper-left network, Fig. 5). When these samples are mapped onto low EF networks, between-node correlation is lost, as reflected by greater color dissimilarity between node pairs (e.g. upper-right network, Fig. 5). An inverted relation, based on the same interpretation approach, was found in low EF samples. This suggested that one could concentrate on a single (class-specific) network to perform automatic sample classification. To assess the potential of this idea, we focused on the high EF network derived with a $c_T = 0.8$ (left column, Fig. 5). This is also suitable because of the small number of nodes and links.

3.5. Network-based prognostic classification

We then investigated the feasibility of an automated computational approach to classification, which integrates expression and co-expression information extracted from the 4-gene network. This task required finding answers to the following questions: which quantitative measurements may be used to summarize the different patient-specific networks, and can such a quantitative measure be used to differentiate between patient classes?

Here we proposed NAIs as the basis for potential alternative patient classification strategies. The idea is to quantify, for a given patient, the levels of expression and co-expression of the four genes in the context of the high EF network. To implement this idea, different statistic scores that combine the four expression values can be implemented. In this study we report three NAIs, which are defined as follows:

\[
\text{NAI}_1 = M(d_{T+L}, d_{V+D})
\]

\[
\text{NAI}_2 = M(d_{T-L}, d_{V-D})
\]

\[
\text{NAI}_3 = M(d_{M}(T, L), d_{M}(V, D))
\]

with “M” representing the mean operator. In Eq. (1), $d_{T+L}$ and $d_{V+D}$ represent the sum of expression values of TNXB and LTPB4 and the sum of expression values of VEGFB and DAXX, respectively.

Fig. 4. Visualization of molecular mechanisms driving phenotypic divergence after MI: biological processes in which genes are kept in sync (or out of sync) after MI, which leads to either cardiac repair or ventricular dysfunction responses.

Fig. 5. Expression data from high (upper row) and low EF (lower row) patient mapped onto the EF class-specific co-expression networks: The expression levels from a given sample are plotted on the network nodes based on a color scale. Color scale is determined by expression values obtained by microarrays: green (low expression), red (high expression). Linked nodes showing similar colors indicate good coordination between the genes in the mapped sample. Linked nodes showing dissimilar colors indicate poor co-expression in the mapped sample. $c_T = 0.8$ for all the networks. This Fig. illustrates how samples originating from different EF classes produce differential patterns of expression and co-expression on the different class-specific networks. It also suggests that the use of one of the class-specific networks (i.e. the high EF network) can be used to distinguish individuals from both classes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
In Eq. (2), \( d_{T,L} \) represents the difference of the expression values between TNXB and LTPB4, and \( d_{V,D} \) is the difference of the expression values of VEGFB and DAXX. In Eq. (3), \( d_{NV(VD)} \) represents the mean expression value of TNXB and LTPB4, and \( d_{(V,D)} \) is the mean expression value of VEGFB and DAXX. These indices encapsulate quantitative gene pair-wise relationships that are potentially relevant to capture between-class differences at the transcriptional level.

Different classification models, such as logistic regression and support vector machines models, were implemented using these indices as inputs. Fig. 6 and Table S2 in Supplementary Information shows representative results from these models, together with other standard approaches based on gene expression data only. The highest performance was obtained by using (cut-off) values of NAI-1 and NAI-3 independently (leave-one-out cross-validation, LOO-CV, area under the receiving operating curve (AUC) = 0.75). Logistic regression models using those NAIs as single inputs independently reported AUC values of 0.68 and 0.69, respectively. The combination of NAI-1 and NAI-3 into a single logistic regression model produced an AUC = 0.65. A logistic regression model that integrated NAI-1 and NAI-2 reported an AUC = 0.74. Lower prediction performance was obtained with standard classifiers using gene expression data only (linear Support Vector Machine, maximum AUC = 0.69), which was built on nine genes selected by SAM [13]. Classifiers based solely on the expression values of the four genes as input values, i.e. without between-gene information, showed maximum AUC = 0.68 (logistic regression model). Although absolute differences were observed between proposed and standard models, these were not statistically detectable at \( P = 0.05 \). A possible lack of statistical power may be explained by the relatively small number of patients included in this study [23,24]. Thus, future studies will require the analysis of larger population samples.

### 3.6. Independent validation

The obtained NAI-based prediction models were tested on an independent dataset consisting of 62 patients not included above (Methods, Table S3 in Supplementary Information). Clinical characteristics of these patients were undistinguishable to the previous cohort (Table S1 in Supplementary Information) and were divided in two groups based on their EF 1-month after MI, as previously performed: low EF (\(<40\%\); median: 38, range: 17–40) or high EF (>40\%; median: 55, range: 42–86). Classification accuracies were consistent, and in some cases higher, that those shown above (Fig. 6, Table S4 in Supplementary Information). As in the microarray-based model analysis, the use of individual NAIs as model inputs was sufficient to produce relative high classification performance (AUC = 0.75). The combination of NAI-1 and NAI-3 also reported a maximum AUC = 0.75. Moreover, NAI-based models again outperformed models whose inputs represented the gene expression values of the four genes only, i.e. no between-gene expression relationships were included.

As a second model validation strategy, we built and tested models on the independent evaluation dataset through LOO-CV. The classification performance, as estimated by AUC values, was consistent with the other evaluations (Fig. 6, and Table S5 in Supplementary Information). Different models based on NAI-1, NAI-2 and NAI-3 values again reported accurate prognostic capacity (maximum AUC = 0.77).

It is worth noting that, as in the case of microarray data, reductions of gene co-expression were detected in low EF. In high EF samples, the partial correlation between VEGFB and DAXX was equal to 0.30 (\( P < 0.05 \)). In low EF samples, their partial correlation was reduced to 0.19 (\( P > 0.05 \)). With regard to TNXB and LTBP4: in high EF samples, their partial correlation was –0.18 (\( P < 0.05 \)), and in low EF was reduced to –0.35 (\( P > 0.05 \)).

### 4. Discussion

In this study, we integrated and analyzed: (1) a network of interactions between proteins known to be involved in angiogenesis and MI, (2) gene expression data from blood samples of MI patients, and (3) associations between co-expression networks and ventricular dysfunction after MI. These investigations led to the

![Fig. 6](image-url) Representative results from NAI-based models and standard approaches. AUC: area under the receiver operating characteristic curve; mm: models built/tested on microarray data using LOO-CV; ind1: microarray-derived models tested on independent cohort; ind2: models built and tested through leave-one-out cross validation on independent cohort. Models: NAI-1 (cutoff-based prediction using NAI-1); NAI-3 (cutoff-based prediction using NAI-3); NAI-1-LR (logistic regression with NAI-1 as input); NAI-2-LR (logistic regression with NAI-2 as input); NAI-3-LR (logistic regression with NAI-3 as input); NAI-1/2-LR (logistic regression with NAI-1 and NAI-2 as inputs); NAI-1/3-LR (logistic regression with NAI-1 and NAI-3 as inputs); 4-biomarker (logistic regression with expression values of VEGFB, TNXB, LTBP4, DAXX as inputs); Traditional (support vector machine built with nine genes derived from standard data-driven analysis, SAM method).
identification of four genes, biologically meaningful in the context of left-ventricular function, and useful to predict its status after MI.

A PPI network related to angiogenesis and MI was constructed as a knowledge reference to guide the reduction of the biomarker search space. Our main hypothesis was that angiogenesis may offer protection in the setting of myocardial ischemia. The second main assumption was that proteins linked via physical interactions are likely to be functionally-related in specialized processes or functions relevant to angiogenesis and MI. We focused our attention on the network cluster that exhibited a high enrichment of proteins involved in the regulation of angiogenesis. Our hypothesis was that this cluster could contain primary drivers of angiogenesis in the context of MI. We went beyond the analysis of well-characterized angiogenesis-related genes to search for potentially novel biomarkers linked to both angiogenesis and ventricular dysfunction.

Matching of co-expression networks to EF classes, i.e. to preserved LV function (high EF) or failure (low EF), identified four main candidates, VEGFB, DAXX, TNXB and LTBP4. These proteins are either involved in angiogenesis, apoptosis or in the processing of the extracellular matrix. VEGFB is a member of the VEGF family of growth factors and its involvement in the recovery of the heart after ischemia has been evidenced using VEGFB-deficient mice [25]. DAXX interacts with several proteins involved in cell death, although its precise role in apoptosis is still unclear [26]. TNXB plays a causative role in Ehlers–Danlos syndrome through regulation of collagen deposition by fibroblasts [27]. LTBP4 is involved in the turnover of the extracellular matrix and in fibroblast adhesion [28].

This investigation offers clinically-relevant insights into the reprogramming of patient-specific transcriptional networks, which may prompt the occurrence of ventricular dysfunction. Co-expression networks characterizing high and low EF samples were identified, which are differentiated by structural and functional features. This allowed visualization of a transcriptional circuitry reconfiguration distinguishing clinical outcomes after MI. Among different responses, activation of collagen metabolic processes, together with major perturbations of cell proliferation and growth regulation, are triggered along the route leading to low EF. Indeed, the most striking difference between high and low EF patients relates to regulation of growth. Whereas high EF patients display a relatively strong co-expression between genes involved in growth regulation, this co-expression seems to be lost in low EF patients. This observation is consistent with cell growth as a natural repair mechanism of the heart after MI. In addition, we noted a significant coordination between genes implicated in collagen metabolism in the low EF group, presumably in relation to an excessive synthesis of collagen which can lead to fibrosis.

The expression relationships between VEGFB, DAXX, TNXB and LTBP4 have prognostic value in the majority of the samples included here. Information extracted from the gene co-expression interplay of these four genes allowed us to build classification models with AUC > 0.70 in different cross-validation and independent evaluation scenarios. The NAI-based models proposed here represent a potentially novel network-based approach to classification, which does not rely on information obtained from individual genes or their expression values independently.

One of the possible limitations of this study is the relatively small amount of samples. However, it should be stressed that our proposal of combining different knowledge and data resources was conceived in response to this reality, as well as the limited power associated with single data sources. Despite the limited amount of data, novel testable insights have been provided regarding the evolution of ventricular dysfunction after MI. We also demonstrated how quantitative and qualitative evidence can be used to design new and relatively powerful prognostic methodologies in a real clinical environment, including the application of cost-effective PCR multiplex assays.

Another critical aspect is the known limits of using protein/gene networks for predictive data mining as argued by Hakes et al. [29]. In this investigation, nevertheless, the PPI represented a knowledge reference to guide subsequent analyses. Such a framework allowed us to focus our attention on a relatively small group of proteins based on their relevance to post-MI processes. Our main hypothesis was that angiogenesis may play a cardioprotective role in the setting of myocardial ischemia. The initial set of 35 genes is known to be involved in angiogenesis and was used as the “seed list” of the PPI network construction. We acknowledge that this may be a relatively small number for knowledge discovery. However, the network construction process expanded the biomarker search space to more than 500 genes. Moreover, the new biomarkers proposed here were not included in the seed list of genes. Other PPI databases could have been analyzed in the implementation of the approach presented here. We chose the HPRD because it has become a well-known resource for supporting network-based knowledge discovery in different biomedical domains. HPRD also offers a relative large number of PPIs obtained from in vivo and in vitro methods. Furthermore, HPRD has become a community-driven database [30], which in addition is currently used by key biological knowledge resources, such as Entrez-Gene, STRING and the MSigDB [30]. The use of multiple databases may have contributed to the representation of a more comprehensive, though not necessarily more “accurate”, network of PPIs related to angiogenesis. However, we aimed to construct the PPI network as a resource to guide the search for potentially relevant biomarkers.

By concentrating on first-interacting partners we may have missed potential biologically-meaningful interactions. Nevertheless, an emphasis on first-interacting partners is also a strategy to reduce the number of potential false positive interactions, i.e. spurious associations in the context of angiogenesis and MI in humans. Our main objective was to focus our attention on proteins and interactions with potential significant roles in angiogenesis after MI. Also, it is important to stress that the different networks reported here were not used to infer topological motifs or unknown relationships with potential functional relevance. We also recognize that it would be possible to propose more sophisticated methods for reconstructing co-expression networks [31]. However, the aim of this study was not to present a detailed view of the transcriptional regulatory architecture in each clinical class. In any case, even if required, such level of resolution would not be possible based on the number of samples available and the lack of time-dependent data. Our main goal was to identify major, global differential patterns between patient groups using co-expression information.

5. Conclusions

In summary, we propose a new strategy for analyzing differential transcriptional responses after MI, and for classifying patients on the basis of both gene expression and co-expression signatures. This proof-of-concept study is compatible with ongoing efforts towards personalized medicine. It may aid clinicians in defining priorities for treatment or prevention of ventricular dysfunction post-MI, and drug developers in identifying participants for clinical trials. Another advantage is the capacity to detect a small number of potential biomarkers, whose integrated expression coordination is altered in patients exhibiting complications after MI.

Competing interests

The authors declare that they have no competing interests.
Acknowledgments

We thank Céline Jeanty, Bernadette Lenars, Malou Gloesener and Loredana Jacobs for expert technical assistance. Funding: This work was supported by grants from the Société pour la Recherche sur les Maladies Cardiovasculaires; Ministère de la Culture, de l’Enseignement Supérieur et de la Recherche, and Fonds National de la Recherche of Luxembourg.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jbi.2010.05.012.

References