BBADIS-16-507-R1 1 Integrative network analysis reveals time-dependent molecular events underlying left ventricular remodeling in post-myocardial infarction patients

Florence Pinet, Marie Cuvelliez, Thomas Kelder, Philippe Amouyel, Marijana Radonjic, Christophe Bauters

Integrative network analysis reveals time-dependent molecular events underlying left ventricular remodeling in post-myocardial infarction patients

Florence Pinet[1], Marie Cuvelliez[1], Thomas Kelder[2], Philippe Amouyel[3], Marijana Radonjic[2], Christophe Bauters[3]

[1]Inserm, U1167, Univ. Lille, Institut Pasteur de Lille, FHU-REMOD-VHF, F-59000 Lille, France

[2] EdgeLeap B.V., Utrecht, the Netherlands

[3]Univ. Lille, Inserm, U1167, CHRU Lille, Institut Pasteur de Lille, F-59000 Lille, France

Corresponding author:

Dr Florence PINET, INSERM U1167-IPL, 1 rue du professeur Calmette, 59019 Lille cedex, France

Tel: (33) 3 20 87 72 15

Fax: (33) 3 20 87 78 94

E-mail: florence.pinet@pasteur-lille.fr

Word count: 6413 words

Abstract

To elucidate the time-resolved molecular events underlying the LV remodeling (LVR) process, we developed a large-scale network model that integrates the 24 molecular variables (plasma proteins and non-coding RNAs) collected in the REVE-2 study at four time points (baseline, 1 month, 3 months and 1 year) after MI. The REVE-2 network model was built by extending the set of REVE-2 variables with their mechanistic context based on known molecular interactions (1,310 nodes and 8,639 edges). Changes in the molecular variables between the group of patients with high LVR (>20 %) and low LVR (<20 %) were used to identify active network modules within the clusters associated with progression of LVR, enabling assessment of time-resolved molecular changes. Although the majority of molecular changes occur at the baseline, two network modules specifically show an increasing number of active molecules throughout the post-MI follow up: one involved in muscle filament sliding, containing the major troponin forms and tropomyosin proteins, and the other associated with extracellular matrix disassembly, including matrix metalloproteinases, tissue inhibitors of metalloproteinases and laminin proteins. For the first time, integrative network analysis of molecular variables collected in REVE-2 patients with known molecular interactions allows insight into time-dependent mechanisms associated with LVR following MI, linking specific processes with LV structure alteration. In addition, the REVE-2 network model provides a shortlist of prioritized putative novel biomarker candidates for detection of LVR after MI event associated with a high risk of heart failure and is a valuable resource for further hypothesis generation.

Key words: Left ventricule remodeling, echocardiography, biomarkers, system biology

1. Introduction

Left ventricular remodeling (LVR) after myocardial infarction (MI) is a strong predictor of heart failure and cardiovascular death[1]. In humans, imaging studies have confirmed a similar progressive left ventricular dilation after MI[1,2] but difficulties in accessing to myocardial samples in this setting have precluded a comprehensive assessment of the time-dependent changes occurring at the tissue level.

As an alternative to directly studying human myocardial biopsies, circulating biomarkers are easy to assess and may shed light on the pathophysiologic processes involved in heart failure and LVR[3]. Although a reasonable number of circulating biomarkers have been associated with LVR, most studies published so far have focused on one or few markers, and in most cases at a single time-point after MI [reviewed in [4]]. Therefore, the complexity and the time-dependent nature of LVR after MI in humans have not yet been addressed adequately.

It is now recognized that complex biological processes implicated in human diseases are often the result of multiple pathways interacting through interconnected networks that can best be studied using system biology approaches[5,6]. Therefore, in this study we set out to combine experimental data from a serial blood sampling throughout the first year post-MI with information contained in prior knowledge databases to allow more comprehensive, network-based mapping of the dynamic of LVR process in humans.

The aim of the present study was to gain insight into molecular mechanisms that are associated with LVR at different timepoints after MI and to discover putative novel biomarker candidates for detection of LVR after a MI event. To this end, molecular data collected in the REVE-2 study, a prospective cohort of 246 patients dedicated to the study of circulating biomarkers after anterior MI[7], has been integrated with known molecular interactions available in 12 public knowledge databases. The resulting network model has been analyzed to explore processes associated with LVR at different timepoints after MI and to generate hypotheses regarding putative novel biomarker candidates for timely detection of LVR after a MI event. Together, for the first time we gain mechanistic insights in post-MI LVR process in such comprehensive and dynamic way, paving the way for further testing of generated hypotheses.

2. Methods

2.1. The REVE-2 Study

The design of the REVE-2 study has been published in detail elsewhere[7]; this prospective multicenter study was designed to analyze the association between circulating biomarkers and LVR. We enrolled 246 patients with a first anterior wall Q-wave MI. Inclusion criteria were hospitalization within 24 hours after symptom onset and at least 3 akinetic left ventricular segments in the infarct zone at the predischarge echocardiography. Exclusion criteria were inadequate echographic image quality, life-limiting noncardiac disease, significant valvular disease, or prior Q-wave MI. The protocol required serial echographic studies at hospital discharge (baseline = day 3 to day 7), 3 months (3M) and 1 year (1Y) after MI to assess LVR which was defined as a change in left ventricular end-diastolic volume between baseline and 1Y >20%. Serial blood samples were taken at baseline, 1 month (1M), 3M, and 1Y after MI. The institutional Ethics Committee (Centre Hospitalier Universitaire de Lille) approved the study; written informed consent was obtained from all patients.

The characteristics of the 246 patients included in the REVE-2 study are shown in Table 1. One-year echocardiographic follow-up was achieved in 226 (92%) patients; 87 (38.5%) patients had LVR.

2.2. Molecular data processing

The molecular REVE-2 network model was built based on the molecular data [7,8,9,10,11,12,13]. This included 24 molecular variables (18 proteins, 6 non-coding RNAs (5 miRNAs and 1 long non-coding RNA)), measured at 1 to 4 time points (depending on the variable, including baseline, 1M, 3M, and 1Y).

Based on the skewed distributions, the data for all molecular variables were log2 transformed. For variables containing zero or negative values, an offset was applied before transformation by adding the minimum and first non-zero value to the data. To test for significant changes in the measurements between the groups of patients with high LVR versus the group of patients with low LVR, an unpaired, two-sided t-test was performed for each variable at each given time point. The resulting p-values were corrected for multiple testing using the Benjamini-Hochberg FDR control

procedure[14]. Unless specified otherwise, all statistical analyses were performed in R version 3.2.0 (http://www.r-project.org/).

2.3. Building the network model

To construct the REVE-2 network model, the knowledge platform *EdgeBox* (EdgeLeap's proprietary knowledge platform) was used as a resource of public knowledge on molecular interactions. This result in embedding REVE-2 variables into their molecular context based on 12 public databases (Supplemental Table 1): ENCODE (http://encodenets.gersteinlab.org), EnsemblGenes (http://www.ensembl.org), HMDB (http://www.hmdb.ca), Microcosm (http://ebi.ac.uk/enright-srv/microcosm), miRBase (http://mirbase.org), miRecords (http://c1.accurascience.com/miRecords), miRTarBase (http://mirtarbase.mbc.nctu.edu.tw), Reactome (http://www.reactome.org), STRING (http://string-db.org), TargetScan (http://www.targetscan.org), TFe (http://www.cisreg.ca/cgi-bin/tfe/home.pl) and WikiPathways (http://www.wikipathways.org).

The REVE-2 network model embeds the REVE-2 molecular variables in an integrated network of known interactions ("edges") between molecules ("nodes"). To enable linking the variables measured in REVE-2 study to their corresponding molecules in the network model, the variable names were manually annotated to a structured identifiers database. Each variable was mapped to a corresponding entity in Ensembl Genes[15], the Human Metabolome Database[16], and miRBase[17]. In addition, the proteins were annotated to Medical Subject Headings (MeSH)[18], Ensembl Phenotypes (http://www.ensembl.org/Homo_sapiens/Phenotype/All), Disease Ontology[19], and DrugBank[20]. All variables were mapped to a database entity. In case multiple entities were identified for a single variable (e.g. when multiple genes encode for the measured protein), all annotations were included. These annotations were used to build the molecular network model.

Firstly, a set of seed nodes was retrieved from *EdgeBox*, including: 1. All molecule nodes mapping to one or more REVE-2 variables; 2. All molecule nodes that are direct neighbors of the nodes defined in step 1 and 3. All molecule nodes that are part of the shortest paths up to 3 edges between any of the nodes defined in step 1. Here "molecule nodes" are defined as nodes representing

genes, miRNAs and metabolites (type “protein_coding”, “processed_transcript”, “miRNA”, or “metabolite”).

Secondly, all edges between these nodes were queried. If multiple edges between two nodes exist (e.g. multiple sources of evidence for an interaction), they were bundled into a single edge. Furthermore, three quality criteria were applied to filter out low confidence interactions: for edges representing miRNA target interactions, edges were included only when either the target has been validated experimentally (data source is miRTarBase), or the target was predicted in at least three prediction resources (Microcosm, miRecords, and TargetScan); for edges originating from STRING, edges were included only for a STRING confidence score > 800; and for edges originating from WikiPathways, edges with type “in_group” were excluded. The resulting network forms the REVE-2 molecular network model, consisting of 1,310 nodes and 8,369 edges.

The REVE-2 molecular network model is an exploratory model which is not built towards a specific prediction, but is aimed towards uncovering potentially relevant relations (”give me all that is known about”) and generating testable hypotheses based on this model.

2.4. Topology based cluster analysis

Network clustering was performed using the InfoMap algorithm as implemented in the igraph R package (version 0.7.1) (http://igraph.org/r/) for an optimal cluster structure detection in the network. As the InfoMap algorithm assigns each node in the network to a cluster, clusters with less than 5 nodes and clusters with only one or fewer edges per node within the cluster were excluded. The identified clusters were subsequently functionally annotated to biological processes. For each identified cluster an overrepresentation analysis was performed using the Fisher exact test to identify Biological Process terms from the Gene Ontology (GO) that are enriched with genes in the cluster. Significantly enriched GO terms ($p < 0.05$) were added as properties to the network.

2.5. Centrality analysis

For each node in the network the centrality measure “betweenness” was calculated using the igraph R package. The betweenness of a node is the normalized count of shortest paths between all other nodes in the network passing through that node. High betweenness of a node indicates that a

molecule is crucial to maintain functionality and coherence of signaling mechanisms, while a low betweenness indicates a more peripheral role for the molecule.

2.6. Assessment of tissue and/or cell type expression of molecules in the model

Expression for miRNA is based on mammalian miRNA ATLAS [21], expression for proteins is based on the TISSUES database, which combines several resources (Protein Atlas, UniProt, etc.), but excluding text mining based results [22]. The "tissues" column in Supplemental table 4 lists all blood, circulating cells and heart tissues that the miRNA / protein has been identified in.

2.7. Active module analysis

Active modules were extracted from the network model by including all significantly changed REVE-2 molecular variables ($p<0.05$) between high- and low-LVR groups for a given time point, as well as all nodes on shortest paths (up to 3 edges) connecting the significant variables, and direct neighbors of the significant variables. The cluster coverage was calculated as a percentage of molecules of each cluster that are part of active modules.

2.8. Network model visualization

All visualizations of the REVE-2 network model were performed using Cytoscape, version 3.2.1.

3. Results

3.1. The REVE-2 molecular network model

The REVE-2 molecular network model embeds the REVE-2 molecular variables in an integrated network of known molecular interactions and contains 1,310 nodes (1,263 protein-coding nodes, 24 miRNA nodes, 22 metabolite nodes and 1 lncRNA node), and 8,639 edges (Supplemental Table 2).

To obtain a high level overview of processes associated with REVE-2 molecular variables, the network model was clustered based on its topology and resulting clusters were functionally annotated to biological processes. Clustering algorithm identifies groups of nodes that are more highly connected to each other than to other nodes in the network, typically representing biological processes or protein complexes. In total, 33 clusters were identified in the REVE-2 molecular network model. All clusters could be significantly annotated to at least one GO term (Table 2). In total, 20 REVE-2 molecular variables are part of a cluster and 12 clusters contain at least one REVE-2 molecular variable. Table 2 lists the 33 clusters and their corresponding most significantly enriched GO term highlighting different processes important for LVR such as eukaryotic transcriptional regulation (clusters 1, 2 and 5), apoptosis (cluster 14), immune response (cluster 40), creatine and aminoacid metabolic process (cluster 41), signaling pathways (cluster 4, 21, 28 and 29) and extracellular matrix organization (clusters 3 and 9). The REVE-2 molecular network model is visualized in Figure 1. Full network cluster characterization and GO enrichment is provided in Supplemental Table 3.

3.2. Relevance of individual molecules in REVE-2 network model

To assess relevance of individual molecules in the REVE-2 network model, the betweenness centrality was calculated. High betweenness suggests that a molecule is crucial to maintain functionality and coherence of signaling mechanisms and the molecules with the highest betweenness centrality may be relevant as putative biomarker candidates. Supplemental Table 4 lists the nodes with the highest betweenness centrality in the REVE-2 network. This allows identifying the most-central REVE-2 variables (e.g. miR-21-5p, miR-222-3p, miR-423-5p,) as well as novel highly central molecules that were not measured yet but could be good candidates for further investigation as

putative novel biomarkers (hsa-miR-335-5p, EP300, CTCF). Interestingly, the 20 nodes with the highest betweenness centrality are linked to 8 clusters (1, 2, 3, 5, 10, 12, 13 and 16), with cluster 1 represented by 8 nodes (Histone acetyltransferase p300 (EP300), Myc proto-oncogene protein (MYC), transcriptional repressor CTCF (CTCF), transcription factor Sp1 (SP1), double-strand-break repair protein rad21 homolog (RAD21), transcription initiation factor TFIID subunit 1 (TAF1), proto-oncogene c-Fos (FOS) and CCAAT/enhancer-binding protein beta (CEBPB)). To further determine if the molecules with high betweenness centrality could qualify as putative biomarker candidates we assessed whether these miRNAs and proteins are expressed in blood, circulating cells or heart tissues. Of the top 20 molecules, 17 can be mapped to one or more of these tissues, further narrowing down the list of putative novel biomarker candidates. For example, three miRNAs (hsa-miR-26b-5p, hsa-miR-124-3p, hsa-miR-744-5p) would be interesting candidates for further testing based on their high betweenness centrality, expression in circulating cells and novelty (i.e. not yet measured among REVE-2 variables). The expression of molecules with the highest betweenness centrality in the tissues of interest is provided in the Supplemental Table 4.

To benchmark the relevance of REVE-2 network model in the context of known biomarkers of LVR, the collection of 52 known molecular LVR biomarkers[4] has been cross-referenced to REVE-2 molecular network model. In total, of 48 biomarkers which could be annotated to either Ensembl Gene or HMDB identifier, 22 were present in the REVE-2 network model.

3.3. Time-resolved activation of network modules associated with LVR progression

The REVE-2 network model is based on prior knowledge of molecular interactions and as such provides insight into molecular mechanisms associated with variables measured in REVE-2 study, and identifies key molecules that may be relevant in the context of LVR. To assess which of these mechanisms change under LVR conditions in REVE-2 study, and which mechanisms are relevant at which time point after MI, an active module analysis was performed (see methods). The active modules were extracted from the REVE-2 network model for each time point (baseline, 1M, 3M, 1Y), enabling insight into time-resolved molecular changes associated with LVR progression. Each active module highlights the part of the network model associated with REVE-2 molecules

which are differentially expressed between high- and low-LVR groups at each time point (supplemental Table 5). These active modules represent a robust mechanistic context of changed variables and provide insight into molecular interactions and mechanisms relevant at a given time point during LVR progression. Figure 2 shows the active modules for each time point.

The most apparent observation is that the highest number of active modules is found at the baseline and 3M time points. This is reflected in both high number of clusters, as well as in total number of molecules represented in active modules at these time points (Figures 2 and 3). Specifically, while all the clusters in the REVE-2 model (n=33) are represented in active modules at baseline, there is a strong decrease of the number of clusters (n=9) in active modules at 1M. At time point 3M, again a large number of clusters (n=30) are in active modules and finally, only 3 clusters are in active modules at 1Y. Although the number of clusters highlighted at 3M (n=30) was very similar to those at baseline (n=33), a higher ratio of molecules belonging to active modules was detected at baseline (936, corresponding to 71% of the molecules of the REVE-2 network model) compared to 3M (389 molecules corresponding to 30% of the molecules of the REVE-2 network). Figure 3 details cluster coverage of the active modules at each time point.

In contrast to the general pattern of largest cluster coverage in the active module at baseline, clusters 7 and 9 involved respectively in muscle filament sliding and extracellular matrix organization, remain highly represented in active modules at 1M, with the highest cluster coverage at 3M. This pattern suggests that these two clusters and associated processes specifically represent the dynamic process of long-term pathological changes during the post-MI LVR, rather than the acute stress of the system triggered by MI (Figure 3 and Table 3). Cluster 7, which includes proteins of myofilament such as the TnT, TnC and TnI forms, myosin and tropomyosin proteins, has 33 active molecules at 1M and 3M representing the highest number of molecules, 94% compared to only 11% at baseline. Cluster 9 which includes most of the metalloproteinase (MMPs), tissue inhibitors of metalloproteinases (TIMPs) and laminin proteins has a ratio of nearly 60% of active molecules (n=28) at 3M compared to baseline (n=17) and 1M (n=7).

4. Discussion

Due to the difficulties in accessing human myocardial samples after MI, little is known about long-term processes associated with LVR in humans. Studies so far focused on a handful of measurable variables, typically assessed only at the baseline[4]. Our objective was therefore to build a more comprehensive view on processes associated with LVR up to one year after MI, by enriching our dataset obtained from serial blood samples during the LVR progression (baseline, 1M, 3M, 1Y) with prior knowledge on molecular interactions. Network analysis allowed us to identify robust mechanisms associated with the measured variables, as this approach considers not only list of isolated molecules, but also accounts for their mechanistic context. Importantly, availability of time-resolved dataset allowed us to determine relevance of LVR-associated mechanisms over time after an MI event. Interestingly, muscle filament sliding and extracellular matrix disassembly have been identified to be specifically (unlike other identified mechanisms) more activated at the 3M timepoint compared to the baseline, suggesting their relevance for the long-term pathological changes during the post-MI LVR. To the best of our knowledge, here we for the first time can discriminate between mechanisms associated with acute stress response triggered by MI and the mechanisms associated with dynamic process of long-term pathological changes during the post-MI LVR.

Another important result of our study is the prioritization of putative LVR-relevant molecules, including those that have not yet been measured or previously associated with LVR. The undertaken knowledge-based network modelling approach therefore stimulates novel discovery and hypothesis forming, which is particularly relevant for identification of putative novel biomarker candidates for timely detection of LVR after a MI event. In line with this motivation, we provide a shortlist of putative novel biomarker candidates prioritized based on their high betweeness centrality and expression in accessible tissues, which may be considered for a follow up biomarker studies. Circulating biomarkers may provide information on changes occurring in the myocardium after MI, specifically in humans due to the difficulties in accessing to human myocardial samples after MI. MMPs, TIMPs, and cytokines are known to be implicated in the myocardium during the remodeling process[23,24], and are also modulated in a similar time-dependent fashion in the circulation [10,25,26]. In a 2012 review of the literature, 59 studies examining in total 112 relations between 52 different

biomarkers and LVR were cited[4]. Therefore - although this approach can certainly not be viewed as equivalent to direct studies on myocardial tissue, the assessment of circulating biomarkers may provide useful information on LVR and heart failure.

4.1. The REVE-2 network model

Systems and network biology approaches are being applied with the aim to capture the molecular complexity of the disease by data integration, ultimately achieving goals of systems medicine[27]. Integrative approaches have been classified into four broad categories: identification of active modules through integration of networks and molecule profiles, identification of conserved modules across multiple species, identification of differential network modules through integration of different interaction types, and identification of composite functional modules through integration of different interaction types[6]. Recently, a MI-specific protein-protein-interaction (MIPIN) network has been built using 38 seed proteins from published human studies[28]. Based on statistical analysis of individual molecules, the MIPIN network was constituted of 613 proteins that associated 4443 edges interactions. In the present study, we used the molecular variables measured in the REVE-2 study, their direct neighbors and molecules on the shortest paths between these molecules as the seed nodes for building a network model. The interactions between molecules were derived from 12 public resources, including functional links such as physical interaction, pathway interactions, transcriptional regulation and enzyme activation. The resulting network model constitutes 1,310 nodes, including not only 1,263 proteins but also 24 miRNAs, 22 metabolites and 1 lncRNA, connected through 8639 edges. A network topology of clustering the REVE-2 network model identified 33 clusters annotated to Gene Ontology processes. Therefore, the REVE-2 network model provides robust insight in LVR-associated physiopathological mechanisms and facilitates identification of novel potentially relevant molecules to be studied further, which is as mentioned above, an important resource for discovery of novel putative biomarkers. For instance, betweenness centrality of the nodes indicates relevance of individual molecules within the model, where high betweenness suggests that a molecule is crucial to maintain functionality and coherence of signaling mechanisms. Interestingly, we identified as highly central molecules 1 miRNA, miR-335-5p and 2 proteins, the histone acetylase p300 (EP300) and the

transcriptional repressor CCCTC-binding factor (CTCF). The latter proteins have not yet been associated with LVR but they have been described indirectly to have a potential role in the heart. SIRT-1, which is acetylated by EP300 has been described to be implicated in dystrophic cardiomyopathy[29] and a CTCF binding site has been described to be a novel heart enhancer[30]. Importantly, most of the molecules with high betweenness centrality are expressed in blood, circulating cells or heart tissues, further qualifying them for the follow up biomarker studies. For example, three miRNAs (hsa-miR-26b-5p, hsa-miR-124-3p, hsa-miR-744-5p) would be interesting candidates for further testing based on their high expression in circulating cells and novelty.

4.2. *Temporal changes associated with LVR*

Strength of the present study is the availability of serial blood samples during the LVR progression (baseline, 1M, 3M, 1Y). Imaging studies in humans have shown that LV dilation begins early after MI and may continue up to 1 year after the initial insult[2,31]. In animal models, the cellular and molecular changes that occur within the myocardium can last weeks or months after MI [23,24]. The baseline sample (3-7 days) likely combines information on initial myocardial necrosis and on the early steps of LVR. Later samples (1M to 1Y) are at a significant distance from initial MI and are probably more indicative of the late remodeling process.

Our data summarized in Figures 2 and 3 show the active modules of the REVE-2 network model associated with each time point, enabling insight into time-dependent molecular changes associated with LVR progression. Examination of temporal molecular patterns reveals a clear dynamic signature of biological processes. Most of the LVR-relevant processes are activated at the baseline and include molecules which have been previously reviewed[3] as potential heart failure biomarkers such as markers of inflammation and oxidative stress, extracellular matrix remodeling, and myocyte injury. After the initial changes at baseline, at 1M the changes in processes underlying LVR are less prominent, and are reactivated at 3M (although including a smaller subset of molecules in these processes). Interestingly, as outlined above, large number of molecules from the clusters associated with muscle filament sliding and extracellular matrix disassembly, represented in active modules at 3 months confirms a specific role for these processes in the later stage of LVR. The statistical

significance of changes between high- and low-LVR group in REVE-2 molecules associated with these clusters (troponin and MMPs) is in line with this hypothesis, with increase in changes from baseline up to 3M (supplemental Table 5). Finally, the low number of active molecules and clusters detected at 1Y indicates a shut-down of all processes associated to LVR. In summary, our results suggest that many molecular events and pathways are activated during LVR in a time-dependent manner and allow better understanding of which mechanisms are linked to acute stress of the system upon MI, and which are associated with long-term pathological changes associated with LVR.

Conclusion

In conclusion, our results improve our knowledge about the LVR process after MI by identifying the temporal activation profiles of REVE-2 molecules along with their associated mechanistic context. This provides more robust insight in LVR processes compared to purely statistical methods based on limited set of variables, facilitates mechanistic hypotheses and prioritizes molecules to be further studied as putative biomarker for timely detection of LVR after MI. The REVE-2 network model is a promising tool to capture the molecular complexity of the disease and can be used as a resource for discovery of LVR-relevant interactions in the context of development of therapeutic interventions aiming at reducing adverse LVR (e.g., which target? at which time point after MI?). Finally, another implication of interest would be to explore whether the same type of proteins (nodes) segregate into close neighborhood of proteins associated for other diseases, providing clues for investigation of comorbidities[32].

Funding Sources

This work was supported by grants from the E.U. FP7 HOMAGE (305507), from CHRU de Lille (Bonus H2013) and from ANR-15-CEA-016/DGOS 15-013.

Disclosure

FP, MC, PA and CB disclose no conflicts of interest.

TK and MR are founders and owners of EdgeLeap B.V.

Author Contributions

FP, CB conceived and designed the experiments; FP, MC, CB acquired the data; FP, CB, TK, MR designed and executed analysis workflows and interpreted the data; FP, MC, TK, MR, CB drafted the manuscript; FP, TK, PA, MR, CB revised the manuscript for important intellectual content; FP, PA, CB obtained funding; FP, CB supervised the study.

References

1. Sutton MSJ, Pfeffer MA, Plappert T, Rouleau JL, Moye LA, Dagenais GR, Lamas GA, Klein M, Sussex B, Goldman S, Menapace Jr FJ, Parker JO, Lewis S, Sestier F, Gordon DF, McEwan P, Bernstein V, Braunwald E, Lewis JO, Sestier F, Gordon DF, McEwan P, Bernstein V, Braunwald E, St John Sutton M, Pfeffer MA, Plappert T, Rouleau JL, Moye LA, Dagenais GR, Lamas GA, Klein M, Sussex B, Goldman S, et al. Quantitative two-dimensional echocardiographic measurements are major predictors of adverse cardiovascular events after acute myocardial infarction. The protective effects of captopril. *Circulation*. 1994;89:68–75.
2. Savoye C, Equine O, Tricot O, Nugue O, Segrestin B, Sauti??re K, Elkohen M, Pretorian EM, Taghipour K, Philias A, Aum??geat V, Decoulx E, Ennezat P V., Bauters C. Left Ventricular Remodeling After Anterior Wall Acute Myocardial Infarction in Modern Clinical Practice (from the REmodelage VEntriculaire [REVE] Study Group). *Am J Cardiol*. 2006;98:1144–1149.
3. Braunwald E. Biomarkers in heart failure. Preface. *N Engl J Med*. 2008;5:xiii–xiv.
4. Fertin M, Dubois E, Belliard A, Amouyel P, Pinet F, Bauters C. Usefulness of circulating biomarkers for the prediction of left ventricular remodeling after myocardial infarction. *Am J Cardiol*. 2012;110:277–83.
5. Azuaje FJ, Dewey FE, Brutsaert DL, Devaux Y, Ashley E a., Wagner DR. Systems-based approaches to cardiovascular biomarker discovery. *Circ Cardiovasc Genet*. 2012;5:360–367.
6. Mitra K, Carvunis AR, Ramesh SK IT. integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*. 2013;14:719–732.
7. Fertin M, Hennache B, Hamon M, Ennezat P V., Biausque F, Elkohen M, Nugue O, Tricot O, Lamblin N, Pinet F, Bauters C. Usefulness of Serial Assessment of B-Type Natriuretic Peptide, Troponin I, and C-Reactive Protein to Predict Left Ventricular Remodeling After Acute Myocardial Infarction (from the REVE-2 Study). *Am J Cardiol*. 2010;106:1410–1416.
8. Lamblin N, Bauters A, Fertin M, De Groote P, Pinet F, Bauters C. Circulating levels of hepatocyte growth factor and left ventricular remodelling after acute myocardial infarction (from the REVE-2 study). *Eur Hear J Hear Fail*. 2011;13:1314–1322.

9. Fertin M, Bauters A, Pinet F, Bauters C. Circulating levels of soluble Fas ligand and left ventricular remodeling after acute myocardial infarction (from the REVE-2 study). *J Cardiol*. 2012;60:93–97.

10. Fertin M, Lemesle G, Turkieh A, Beseme O, Chwastyniak M, Amouyel P, Bauters C, Pinet F. Serum MMP-8: A Novel Indicator of Left Ventricular Remodeling and Cardiac Outcome in Patients after Acute Myocardial Infarction. *PLoS One*. 2013;8:1–6.

11. Bauters C, Kumarswamy R, Holzmann A, Bretthauer J, Anker SD, Pinet F, Thum T. Circulating miR-133a and miR-423-5p fail as biomarkers for left ventricular remodeling after myocardial infarction. *Int J Cardiol*. 2013;168:1837–40.

12. Eschalier R, Fertin M, Fay R, Bauters C, Zannad F, Pinet F, Rossignol P. Extracellular Matrix Turnover Biomarkers Predict Long-Term Left Ventricular Remodeling After Myocardial Infarction: Insights From the REVE-2 Study. *Circ Hear Fail*. 2013;6:1199–1205.

13. Kumarswamy R, Bauters C, Volkmann I, Maury F, Fetisch J, Holzmann A, Lemesle G, de Groote P, Pinet F, Thum T. Circulating Long Noncoding RNA, LIPCAR, Predicts Survival in Patients With Heart Failure. *Circ Res*. 2014;114:1569–1575.

14. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B (Methodol)*. 2009;57:289–300.

15. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M. The Ensembl genome database project. *Nucleic Acids Res* 2002;30:38–41.

16. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013;41:801–807.

17. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34:D140–4.

18. Lipscomb CE. Medical subject Headings (MeSH). *Bull Med Libr Assoc*. 2000;88:265–266.

19. Schriml LM, Arze C, Nadendla S, Chang YWW, Mazaitis M, Felix V, Feng G, Kibbe WA. Disease ontology: A backbone for disease semantic integration. *Nucleic Acids Res*. 2012;40:940–946.

20. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34:D668–D672.

21. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Schliwka J, Fuchs U, Novosel A, Hermida L, Fulci V, Chiaretti S, Foa R, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Russo JJ, Sander C, Zavolan M, Tuschl T. Resource A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. 2007;1401–1414.

22. Santos A, Tsafou K, Stolte C. Comprehensive comparison of large-scale tissue expression datasets. 2015;1–23.

23. Spinale F. Myocardial matrix remodeling and the matrix metalloproteinases: influence on cardiac form and function. *Physiol Rev*. 2007;87:1285–1342.

24. Nian, M, Lee, P, Khaper, N, Liu PP. Inflammatory cytokines and postmyocardial infarction remodeling. *Circ Res*. 2004;94:1543–1553.

25. Webb CS, Bonnema DD, Ahmed SH, Leonardi AH, McClure CD, Clark LL, Stroud RE, Corn WC, Finklea L, Zile MR SF. Specific temporal profile of matrix metalloproteinase release occurs in patients after myocardial infarction: relation to left ventricular remodeling. *Circulation*. 2006;114:1020–1007.

26. Orn S, Breland UM, Mollnes TE, Manhenke C, Dickstein K, Aukrust P UT. The chemokine network in relation to infarct size and left ventricular remodeling following acute myocardial infarction. *Am J Cardiol*. 2009;104:1179–1183.

27. Chan SY, Loscalzo J. The Emerging Paradigm of Network Medicine in the Study of Human Disease. *Circ Res* [Internet]. 2012;111:359–374. Available from: http://circres.ahajournals.org/cgi/doi/10.1161/CIRCRESAHA.111.258541

28. Nguyen NT, Zhang X, Wu C, Lange RA, Chilton RJ, Lindsey ML, Jin Y-F. Integrative computational and experimental approaches to establish a post-myocardial infarction knowledge map. *PLoS Comput Biol*. 2014;10:e1003472.

29. Kuno, A, Tanno, M, Horio Y. The effects of resveratrol and SIRT1 activation on dystrophic cardiomyopathy. *Annu New York Acad Sci*. 2015;1348:46–54.

30. Schultz B, Gallicio G, Cesaroni M, Lupey L, Engel N. Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. *Nucleic Acids Res*. 2015;43:745–759.

31. Gaudron P, Eilles C, Kugler I, Ertl G. Progressive left ventricular dysfunction and remodeling after myocardial infarction. Potential mechanisms and early predictors. *Circulation*. 1993;87:755–763.

32. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J BA. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science (80-)*. 2015;347:1257601.

Table 1. Characteristics of the patients included in the REVE-2 study[1]

Age, y (mean±SD)	57±14
Women	46 (19%)
Diabetes mellitus	51 (21%)
First anterior myocardial infarction	246 (100%)
Initial reperfusion therapy	
- Primary percutaneous coronary intervention	128 (52%)
- Thrombolysis alone	28 (11%)
- Thrombolysis and rescue percutaneous coronary intervention	59 (24%)
- No reperfusion	31 (13%)
Left ventricular ejection fraction, % (mean±SD)	49±8
Medications at discharge	
- Antiplatelet therapy	246 (100%)
- ß-blockers	238 (97%)
- Angiotensin-converting enzyme inhibitors	238 (97%)
- Statins	231 (94%)
One-year echocardiographic follow-up	
- No. of patients with follow-up	226 (92%)
- Left ventricular remodeling[2]	87 (38.5%)

[1] Of the 226 patients with echocardiographic follow-up; [2]Defined as a >20% change in left ventricular end-diastolic volume between baseline and 1 year.

Table 2. List of the identified clusters of the REVE-2 network model.

Cluster number	Nodes[1]	Edges[2]	REVE-2 variables[3]	Most significant GO category[4]
1	191	1013	ST2	sequence-specific DNA binding
2	203	302	Mir-21-5p	intrinsic component of membrane
3	57	978	ICTP, P1NP, P3NP	endoplasmic reticulum lumen
4	65	403	HGF	transmembrane receptor protein tyrosine kinase signaling pathway
5	98	123	miR-222-3p	RNA binding
6	37	495	TIMP1	platelet alpha granule lumen
7	35	392	troponin	muscle filament sliding
8	30	271		nuclear-transcribed mRNA catabolic process
9	47	154	MMP1, MMP2, MMP3, MMP8, MMP9, TIMP2, TIMP4	extracellular matrix disassembly
10	38	75		ATP binding
12	30	75		mitotic cell cycle phase transition
13	32	33		O-glycan processing
14	22	111	FasL	activation of cysteine-type endopeptidase activity involved in apoptotic process
15	17	67		G2/M transition of mitotic cell cycle
16	23	33		G-protein coupled receptor signaling pathway
18	14	61		RNA splicing, via transesterification reactions
19	16	29		positive regulation of intrinsic apoptotic signaling pathway
20	22	23		phosphatase activity
21	14	32		toll-like receptor 2 signaling pathway
22	15	17		steroid hormone receptor activity
24	11	19		regulation of cell differentiation
25	10	18		cell surface
26	11	16		cellular component disassembly involved in execution phase of apoptosis
28	9	14		negative regulation of transforming growth factor beta receptor signaling pathway
29	8	15		small GTPase mediated signal transduction
30	11	12	BNP	protein targeting to mitochondrion
32	9	12		DNA repair
33	9	10		nuclear pore
35	7	11		Mitochondrial inner membrane
37	7	8		DNA-directed RNA polymerase II, holoenzyme
39	6	11		posttranscriptional gene silencing
40	7	12	CRP	complement activation
41	6	11	Creatine kinase	creatine metabolic process

[1]Nodes represent molecules such as proteins, non-coding RNAs and metabolites; [2]Edges represent known interactions between molecules (nodes); [3]REVE-2 variables represent molecules measured in

plasma of REVE-2 patients; [4]Full network cluster characterization and GO (Gene Ontology) is provided in supplemental Table 3.

Table 3. Detailed list of molecules in two clusters modulated during the LVR progression.

Clusters[1]/REVE2 variables[2] (number of nodes)[3]	**Baseline**	**1 month**	**3 months**	**1 year**
7/Troponin (35 nodes)	**Dystrophin (P11532)** **Tropomyosin α-1 chain (P09493)** Caldesmon 1 (Q05682) Myosin 1 (P12882)	**Dystrophin (P11532)** **Tropomyosin α-1 chain (P09493)** TnT, slow skeletal (P13805) TnT, cardiac (P45379) TnT, fast skeletal (P45378) TnI, slow skeletal (P19237) TnI, fast skeletal (P48788) TnI, cardiac (P19429) TnC, slow skeletal and cardiac (P63316) TnC, fast skeletal (P02585) Tropomyosin β chain (P07951) Tropomyosin α-3 chain (P06753) Desmin (P17661) MYBPC1 protein (Q00872) MYBPC2 protein (Q14324) MYBPC3 protein (Q14896) MLC1/3, skeletal (P05976) MLC2, cardiac (P10916) MLC3 (P08590) MLC4 (P12829) Myosin 3 (P11055) Myosin 6 (P13533) Myosin 8 (P13535) Vimentin (P08670) Nebulin (P20929) Tropomodulin-1 (P28289)	**Dystrophin (P11532)** **Tropomyosin α-1 chain (P09493)** TnT, slow skeletal (P13805) TnT, cardiac (P45379) TnT, fast skeletal (P45378) TnI, slow skeletal (P19237) TnI, fast skeletal (P48788) TnI, cardiac (P19429) TnC, slow skeletal and cardiac (P63316) TnC, fast skeletal (P02585) Tropomyosin β chain (P07951) Tropomyosin α-3 chain (P06753) Desmin (P17661) MYBPC1 protein (Q00872) MYBPC2 protein (Q14324) MYBPC3 protein (Q14896) MLC1/3, skeletal (P05976) MLC2, cardiac (P10916) MLC3 (P08590) MLC4 (P12829) Myosin 3 (P11055) Myosin 6 (P13533) Myosin 8 (P13535) Vimentin (P08670) Nebulin (P20929) Tropomodulin-1 (P28289)	-

		Calcipressin-3 (Q9UKAB) Telothonin (O15273) Felodipine Trifluoperazine Levosimendan Dihydroxyaluminium	Calcipressin-3 (Q9UKAB) Telothonin (O15273) Felodipine Trifluoperazine Levosimendan Dihydroxyaluminium	
9/ MMP3, MMP8, MMP9, TIMP2, TIMP4 **(47 nodes)**	**Brevican (Q96GW7)** **Neutrophil elastase (P08246)** **MMP8 (P22894)** **PRSS1 protein (Q3SY19)** **Kallikrein -2 (P20151)** **Marimastat** MMP1 (P03956) MMP2 (P08253) TIMP2 (P16035) MMP9 (P14780) Aggrecan (P16112) hsa-miR-338-3p hsa-miR-451a Captopril HS3SB protein (Q9Y662) Fibrillin-1 (P35555) Nidogen-1 (P14543)	**Brevican (Q96GW7)** **Neutrophil elastase (P08246)** **MMP8 (P22894)** **PRSS1 protein (Q3SY19)** **Kallikrein -2 (P20151)** **Marimastat** Tenascin C (P24821)	**Brevican (Q96GW7)** **Neutrophil elastase (P08246)** **MMP8 (P22894)** **PRSS1 protein (Q3SY19)** **Kallikrein -2 (P20151)** **Marimastat** MMP1 (P03956) MMP2 (P08253) TIMP2 (P16035) MMP3 (P08254) Aggrecan (P16112) hsa-miR-338-3p hsa-miR-451a Captopril Tenascin C (P24821) MMP7 (P09237) MMP11 (P24347) MMP14 (P50281) TIMP4 (Q99727) Decorin (P07585) Laminin alpha3 (Q16787) Laminin beta-3 (Q13751) Laminin gamma2 (Q13753) Syndecan-2 (P34741) IGF2 (P01344) IGFALs protein (P35858) IBP3 protein (P17936) Halofuginone	-

[1]Number corresponds to the cluster listed in Table 2; [2]REVE-2 variables measured in plasma of REVE-2 patients; [3]Number of nodes (molecules) present within the cluster; Genes in bold indicates the molecules present in active modules at each time point; Accession number in UniProtKB (http://www.uniprot.org/) is provided for proteins under bracket.

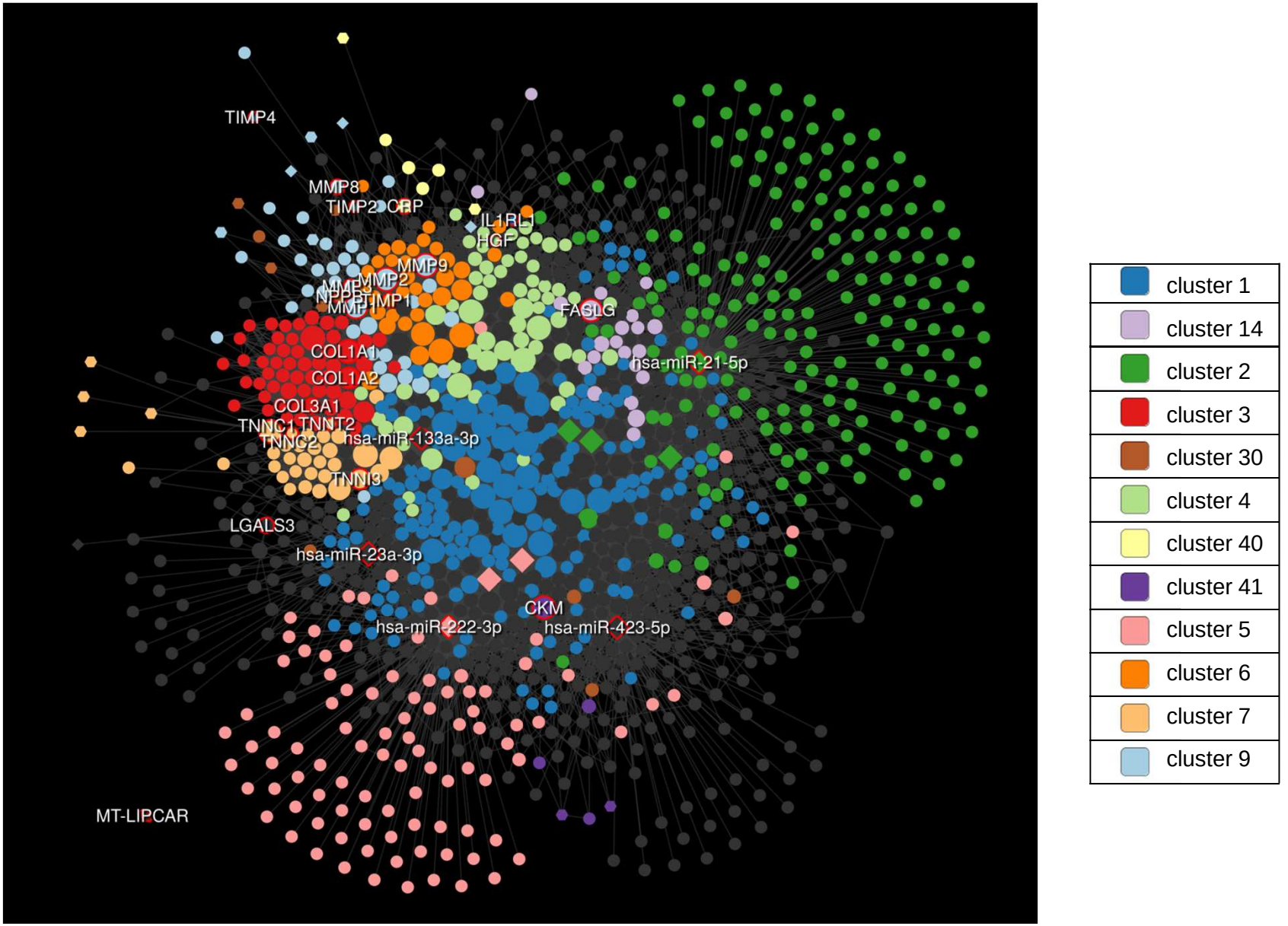
TIMP4
MMP8
TIMP2
CRP
IL1RL1
HGF
MMP9
MMP2
MMP3
NPPB
TIMP1
MMP1
FASLG
COL1A1
COL1A2
hsa-miR-21-5p
COL3A1
TNNC1
TNNT2
TNNC2
hsa-miR-133a-3p
TNNI3
LGALS3
hsa-miR-23a-3p
CKM
hsa-miR-222-3p
hsa-miR-423-5p
MT-LIPCAR
cluster 1
cluster 14
cluster 2
cluster 3
cluster 30
cluster 4
cluster 40
cluster 41
cluster 5
cluster 6
cluster 7
cluster 9

Figure 1

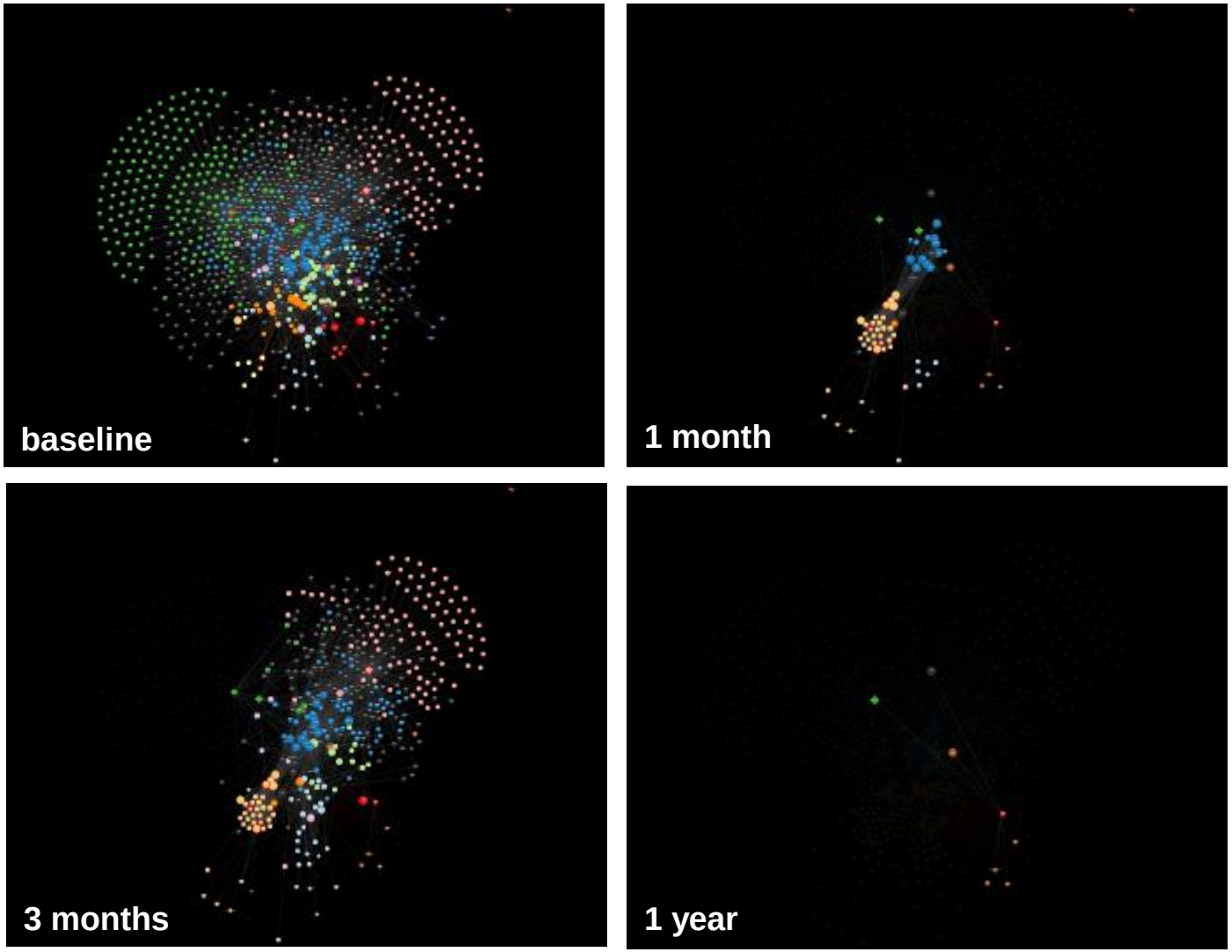

Figure 2

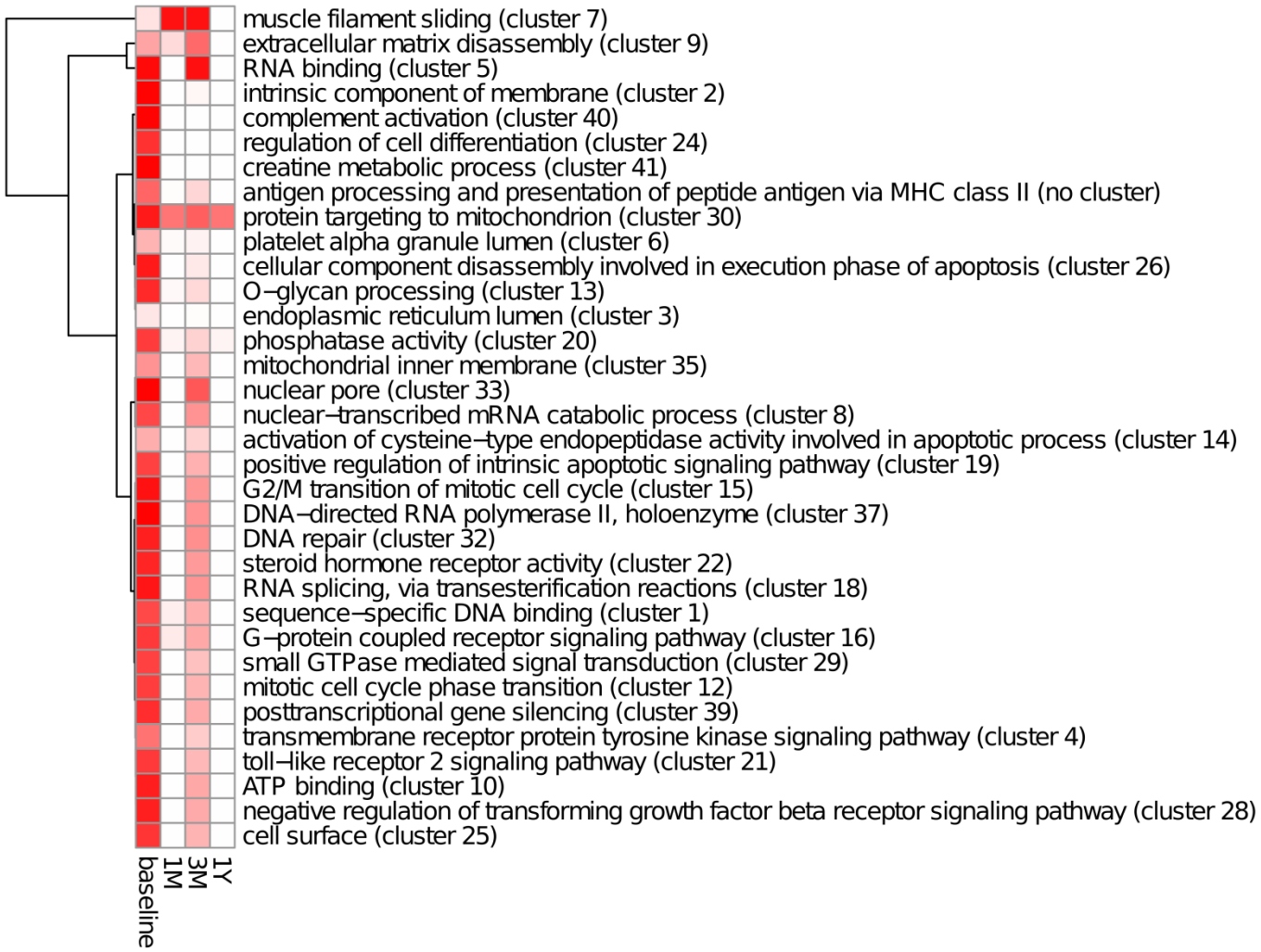
muscle filament sliding (cluster 7)
extracellular matrix disassembly (cluster 9)
RNA binding (cluster 5)
intrinsic component of membrane (cluster 2)
complement activation (cluster 40)
regulation of cell differentiation (cluster 24)
creatine metabolic process (cluster 41)
antigen processing and presentation of peptide antigen via MHC class II (no cluster)
protein targeting to mitochondrion (cluster 30)
platelet alpha granule lumen (cluster 6)
cellular component disassembly involved in execution phase of apoptosis (cluster 26)
O−glycan processing (cluster 13)
endoplasmic reticulum lumen (cluster 3)
phosphatase activity (cluster 20)
mitochondrial inner membrane (cluster 35)
nuclear pore (cluster 33)
nuclear−transcribed mRNA catabolic process (cluster 8)
activation of cysteine−type endopeptidase activity involved in apoptotic process (cluster 14)
positive regulation of intrinsic apoptotic signaling pathway (cluster 19)
G2/M transition of mitotic cell cycle (cluster 15)
DNA−directed RNA polymerase II, holoenzyme (cluster 37)
DNA repair (cluster 32)
steroid hormone receptor activity (cluster 22)
RNA splicing, via transesterification reactions (cluster 18)
sequence−specific DNA binding (cluster 1)
G−protein coupled receptor signaling pathway (cluster 16)
small GTPase mediated signal transduction (cluster 29)
mitotic cell cycle phase transition (cluster 12)
posttranscriptional gene silencing (cluster 39)
transmembrane receptor protein tyrosine kinase signaling pathway (cluster 4)
toll−like receptor 2 signaling pathway (cluster 21)
ATP binding (cluster 10)
negative regulation of transforming growth factor beta receptor signaling pathway (cluster 28)
cell surface (cluster 25)
baseline
1M
3M
1Y

Figure 3