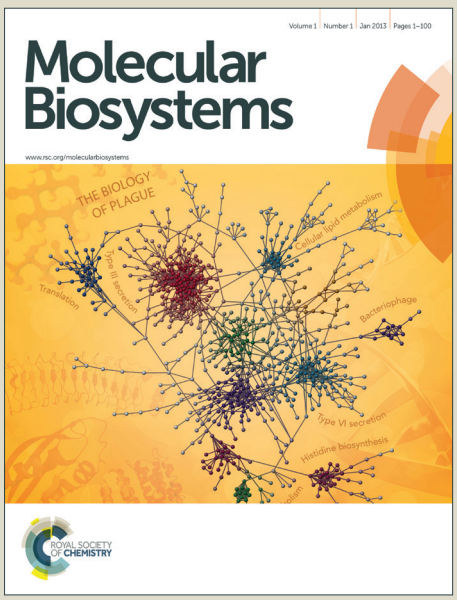
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Robust coordination of cardiac functions from gene co-expression reveals a versatile combinatorial transcriptional control

Catherine Cerutti[1], Giampiero Bricca[1], Sophie Rome[3], Christian Z. Paultre[1,2], and Marie-Paule Gustin[1,2]

[1]EA 4173; Université de Lyon, Université Lyon 1, Hôpital Nord-Ouest Villefranche-sur-Saône, 8 avenue Rockefeller, F-69373, Lyon, France

[2]Université de Lyon, Université Lyon 1, Institut des Sciences Pharmaceutiques et Biologiques (ISPB), 8 avenue Rockefeller, F-69373, Lyon, France

[3]INSERM U1060/INRA 1235; Université de Lyon, INSA Lyon, Université Lyon 1, F-69600, Oullins, France.

Corresponding author:

Catherine Cerutti

Université Lyon 1, EA4173 Génomique fonctionnelle de l'hypertension artérielle

8 avenue Rockefeller

69373 Lyon Cedex 08

France

Tel: +33-426 688 222

Fax: +33-478 777 517

Email: cerutti@univ-lyon1.fr

Abstract

The necessary overall coordination of cardiac cellular functions is little known at the mRNA level. Focusing on energy production and cardiac contraction, we analyzed microarray data from heart tissue obtained in groups of mice and rats in normal conditions and with left ventricular dysfunction. In each group, we identified genes positively or negatively correlated with more genes of one function than random, which were called coordinated or inversely coordinated with the function. The genes coordinated with energy production or cardiac contraction showed the coupling of these functions in all groups. Among coordinated or inversely coordinated genes common to the two functions, we proposed a fair number of transcriptional regulators as potential determinants of the energy production and cardiac contraction coupling. Although this coupling was constant across the groups and unveiled a stable gene core, the combinations of transcriptional regulators were very different between the groups, including one half that has never been linked to heart function.

These results highlighted the stable coordination of energy production or cardiac contraction at the mRNA level, and the combinatorial and versatile nature of potential transcriptional regulation. In addition, this work unveiled new transcriptional regulators potentially involved in the normal or altered cardiac functional coupling.

Keywords: emergence; heart; left ventricular dysfunction; microarray; systems biology; transcriptional regulation

Introduction

Although providing only a picture of the genome at the mRNA level, microarrays represent a precious tool allowing global quantitative studies of the complete genome. In particular one can access cellular functions that are actually complex systems including numerous components or genes, linked together by multiple interactions. Therefore, the behavior of cellular functions cannot consist merely of the sum of individual behaviors of each of its components, and needs global approaches [1–3].

Examining gene co-expression across a group of observations uncovers emergent properties of the group coming from inter-individual expression variability that complement simple mean expression levels of each gene. In addition, extensive co-expression with a set of genes related to a cellular function reflects coordination that is an emergent property of the function. However, although essential in functional genomics, coordination within a function or between functions has been poorly studied in humans or in mammal models except for oxidative phosphorylation [4,5] or for disease-specific sets of genes [6].

In the cardiovascular field, network analyses have been proposed to elucidate mechanisms involved in cardiovascular diseases [7–10]. Analyzing co-expression from differential or raw expression profiles obtained in various conditions, expression modules could be identified and related to various functions [11,12]. In addition, examining promoter regions of the genes within modules may allow identifying some potentially associated transcription factors [13]. However, coordination between modules, which is a necessary feature to ensure efficient functional cooperation, has not been yet sufficiently explored.

From a methodological point of view, most computational methods of co-expression network analysis do not separate the positive and the negative correlations between gene expression profiles or focus only on positive ones [12,14,15]. However network reconstruction tools generally discriminate them for the visualization of the network [16]. Considering that the biological meaning of positive and negative correlations is very different, we previously proposed to manage them separately for a functional interpretation [17].

The aim of this work was to highlight from gene co-expression the coupled coordination of two major cardiac cellular functions, energy production and cardiac contraction, and to identify transcriptional regulators as potential determinants of coordination. We performed a meta-study using already published transcriptomic data obtained in mouse and rat heart tissue. In various conditions, we uncovered a constant coupling of the two functions, which involved large combinations of potential transcriptional regulators. These combinations were very different across the conditions although the gene sets subserving the coupling of the two functions were quite stable. These results highlight the versatility of the transcriptional control across different contexts in contrast with the stability of the functional coupling.

Material and Methods

Microarray data

We used already published microarray data obtained from rat and mouse left ventricular tissue samples [18,19]. In order to detect general rules underlying gene co-expression in normal or altered cardiac functioning, we redefined groups gathering together samples from different strains that had been initially processed separately. We could thus take advantage of enlarged expression variability, including both inter-individual variability within one strain and genetic variability between strains.

The rat dataset was previously collected by our team in the context of studies on hypertension. As in humans, the hypertensive models develop naturally left ventricular hypertrophy in association with hypertension inducing progressive cardiac dysfunction towards heart failure or stroke. Microarray data were obtained with Affymetrix RAE230A GeneChips (Affymetrix Inc., Santa Clara, CA, USA), and they are publicly available in the ArrayExpress database (accession number E-MEXP-357) [18]. It was obtained from 16 adult rats belonging to three different models of hypertension (spontaneously hypertensive, Lyon hypertensive, and transgenic heterozygous TGR(mRen2)27+/- rats) developing left ventricular hypertrophy (LVH), and from 16 normotensive rats (CTR) belonging to their respective

normotensive control strains (Wistar-Kyoto, Lyon normotensive, and TGR(mRen2)27-/-) [20,21].

The mouse dataset was obtained from the public Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/, accession number GSE12413). The experiments consisted in a pharmacological alteration of the cardiac function in mice. Data were obtained with Affymetrix Mouse Genome 430 2.0 Array from adult male mice 10 to 12 week-old that belonged to three inbred wild-type strains (C57, FVB, and B6129SF2/J mice). We analyzed data from 34 mice that underwent 14-day infusions of isoproterenol (Isop), a nonspecific β-adrenergic receptor agonist inducing left ventricular dysfunction, and from 31 control mice receiving no treatment (Ctrl) [19]. Isoproterenol-treated mice exhibited left ventricular dysfunction but no ventricular hypertrophy. Therefore the cardiac dysfunction was the common phenotype between LVH rats and Isop mice.

All datasets were normalized with the robust multichip average method [22], and all data were log2-transformed. Expression changes between groups were tested with the Significance Analysis of Microarrays procedure using the R language (http://www.r-project.org/), and statistical significance was taken at the 5% level of the false discovery rate [23].

Functional coordination study

Our method relies on the selection of genes using successive filters as shown on the flow diagram in **Figure 1A**. The probe sets whose mean expression was lower than the first quartile of the general expression distribution over the whole array in each group were excluded from the analysis.

Functional gene sets

We focused on the two major cardiac functions: 1. energy production (EnProd), including fatty acid oxidation, glycolysis, tricarboxylic acid cycle, and oxidative phosphorylation, and 2. cardiac contraction and calcium handling (CC). The related gene sets were downloaded from the Kyoto Encyclopedia Genes and Genomes database for rats and mice separately (http://www.genome.jp/kegg/pathway.html). The codes of the pathway maps were: 00071 (fatty acid oxidation), 00010 (glycolysis), 00020 (tricarboxylic acid cycle), 00190 (oxidative phosphorylation) for EnProd, 04260

(cardiac muscle contraction) and 04020 (calcium signaling pathway) for CC. In rats and mice, probe sets were about 220 for EnProd, and 120 for CC. For each species and each gene set, the detailed lists of genes and their expression in the different groups are given in Supplementary Table 1.

Significant Pearson correlations

Pearson correlation coefficients were computed between two genes – one gene of the microarray and one EnProd or CC gene - across a group of n observations using the R language. The statistical significance of correlation was set at $p<0.05$ after Bonferroni correction. For the mice data, the correlation thresholds were $|r|>0.6416$ and $|r|>0.6175$, for n=31 and n=34 respectively ($p<10^{-4}$ before correction). To ensure robustness of the correlations in samples of smaller size in rats (n=16), we used a Jackknife resampling procedure with a correlation threshold $|r|>0.8368$ (n=15, $p<10^{-4}$ before correction) [24]. Gene pairs positively or negatively significantly correlated in more than 80% of the samples were taken as robust.

Significant counts of correlations

We developed an original statistical procedure to define coordination with a functional gene set F. For each gene of the microarray, its positive and negative significant correlations with the functional gene set F were counted. Then we compared these counts to those obtained with 1000 random gene sets of similar size and similar expression obtained from the same data set. We selected the genes that were correlated with significantly ($p<0.05$) more genes of F than of random gene sets and that were not part of F: those positively correlated were called coordinated with F (Coord), and those negatively correlated were called inversely coordinated (InvCoord) with F (**Figure 1B**).

Functions coordinated with F

Analyzing separately Coord and InvCoord genes, we examined their Gene Ontology (GO) annotations to identify the cellular functions coordinated with F. We used the R Bioconductor packages topGO and AnnotationDbi [25]. The GO annotations were dated 2013-09 for both the RAE230a and Mouse430-2 Affymetrix GeneChips. General GO terms involving more than 2000 and 3000 genes in the rat and mouse microarray respectively were not considered. We identified the GO terms significantly

enriched by Coord or InvCoord genes (false discovery rate < 5%).

Potential determinants of coordination

We focused on the Coord and InvCoord genes of F known as transcriptional regulators (TR), either DNA-binding transcription factors, transcription cofactors or chromatin modifiers. This stage of the analysis was based on the accepted notion that co-expression of TR with their targets is most often high [26]. The lists of TR were obtained from the MatBase module of Genomatix Software Suite (Genomatix Software GmbH, Munich, Germany) [27]. The underlying database contains information on TR for about 100 different species, coming with references to scientific publications. The lists of TR included 1074 and 1534 probe set IDs in rat and mouse microarray respectively.

Coupled coordination of two functions

Considering two functions F1 and F2, the TR Coord or InvCoord with each function defined four situations according to the sign of coordination with F1 and F2: TR++, TR--, TR+-, and TR-+ (**Figure 1C**)**.** Due to the mathematical transitivity of the correlation, TR++ and TR-- are normally associated with the same F1 and F2 genes that should vary in the same direction reflecting coupled coordination of F1 and F2. Differently, TR+- and TR-+ are likely associated with other F1 and F2 genes that should inversely vary reflecting coupled inverse coordination of F1 and F2. The four types of TR were selected from our data.

Bibliographical analysis

A thorough bibliographical analysis of the selected TR helped detailing their known role in heart or muscle function or development, energy metabolism, inflammation, general development or in other function excluding cancer.

Potential transcription factors

For each DNA-binding transcription factor among the selected TR, the over-representation of binding sites in the gene promoters of each functional gene set was tested with the Gene2Promoter and CommonTFs Genomatix modules (significance taken at $p<0.05$). The Genomatix database contains the weight matrices used to locate potential binding sites of transcription factors in DNA sequences.

Results

Expression analysis

All of the genes included in the functional gene sets had high levels of expression in all of the studied groups (Supplementary Table 1). No gene of the microarray had significantly modified expression in LVH compared to CTR rats. This is in accordance with our previous study that had shown only one gene with similar significant expression changes in the three LVH models compared each to their own controls [18]. In contrast, dramatic changes were observed in Isop vs Ctrl mice (5469 over-expressed and 3645 under-expressed). Differentially expressed genes enriched GO terms mostly related to the cardiovascular function, in accordance with the presence of left ventricle systolic dysfunction after isoproterenol infusion in most of the mice [19]. Concerning EnProd and CC gene sets, 25 to 35% of the genes were differentially expressed mixing over- and under-expressed genes.

Because each group gathered together rats or mice from 3 different strains, we checked the significance of expression differences between the 3 strains over the genes of the EnProd and CC functions (337 Probe Set IDs in rats, 335 in mice). We identified a majority of the genes with significant difference over the subgroups in rats (CTR: 63%, LVH: 67%) and in Ctrl mice (70%), whereas only 20% were different in the Isop group.

Coordination with the functional gene sets

For each functional gene set, we first identified all the genes external to the gene set that were positively or negatively correlated to at least one gene of the gene set. In **Figure 2**, the histogram shows that in CTR rats the number of genes having 3 or more positive correlations was significantly higher ($p<0.05$) with EnProd than with a random gene set of similar size. A total of about 500 genes were correlated with 3 or more EnProd genes in this example. In all groups for EnProd and CC, we could detect similar thresholds beyond which a gene could be declared Coord or InvCoord with the gene set, and we selected those that did not belong to the gene set (**Supplementary Table 2**).

For a given group and a given function, the sets of Coord genes and InvCoord genes were distinct

with almost no common ones, showing the consistency of their selection. The numbers of Coord and InvCoord genes were highly variable according to the function and the group, ranging from 67 for Coord genes with CC in Isop mice to 2695 Coord genes with EnProd in Ctrl mice. In addition, Coord genes were more numerous than InvCoord genes (see detailed counts in Supplementary Table 2). Although many Coord genes with EnProd were common to Isop and Ctrl mice (52%), in other cases only 4 to 28% of Coord or InvCoord genes were common to Isop and Ctrl mice or LVH and CTR rats.

Coordination with the CC gene set

Figure 3 illustrates the coordination with the CC gene set in CTR rats, showing the Coord and InvCoord genes correlated with the largest number of CC genes. These links uncovered a subset of eleven CC genes that was markedly involved both in coordination and in inverse coordination. This small emergent core of the CC gene set contains in particular genes coding for cardiac contractile proteins (actin, heavy and light myosin chain, troponin isoforms, and tropomyosin) as well as the Ca++ transporting ATPase of the cardiac muscle (Atp2a2, alias Serca2) and the mitochondrial carrier adenine nucleotide translocator (Slc25a4, alias Ant). Interestingly, many Coord genes were part of the EnProd gene set, showing mutual coordination between CC and EnProd.

Functional annotations of genes coordinated with EnProd and CC gene sets

In rats and in mice, the genes Coord or InvCoord with a functional gene set, but not belonging to it, enriched GO terms that were much less numerous for InvCoord than for Coord genes, even absent in some groups (see detailed lists in Supplementary Table 3). This result shows the scattering of InvCoord genes across many functions. For Coord genes, the enriched GO terms could be quite easily merged into four families of biological processes: energy metabolism, muscle processes, protein processes, and non-muscle development.

Figure 4 shows that genes Coord with CC enriched GO terms related to energy metabolism in all groups, most notably in mice. The most enriched GO terms were: tricarboxylic acid cycle, and electron transport chain or ATP synthesis coupled proton transport, highlighting the coupling between the CC and EnProd gene sets in all groups. Genes Coord with EnProd enriched GO terms

related to the four families. The enrichment of the energy metabolism family enlarged the EnProd gene set to other energy processes. GO terms related to protein processes (from ribosome biogenesis to protein complex assembly) or to development involved more Coord genes in mice than in rats. In all groups, the presence of muscle processes clearly confirmed the coupling between EnProd and CC.

Coupling and inverse coupling of EnProd and CC gene sets: potential determinants

The analysis of the four types of TR Coord or InvCoord with the two gene sets, TR++ or TR--, TR+- or TR-+, identified a total of 57 distinct well-expressed TR (expression > percentile 40) including 4 to 26 ones per group (**Tables 1 and 2**). The TR++ and TR-- were more numerous than the TR+- and TR-+, which were absent in mice. The most striking result was the low number of TR common to several groups, with only 3 TR++ (Ankrd1, Epas1, Fhl2) and 2 TR-+ (Elk3, Meox2) found in a maximum of 2 groups.

TR as potential determinants of EnProd and CC coupling

Table 1 gives the 42 TR++ or TR-- potentially involved in the coupled coordination of EnProd and CC, which were highly expressed (mean expression percentile = 79). Strikingly, 20 TR mostly found in Ctrl mice have not been described in relation with heart or muscle activity or with energy metabolism. Moreover, 9 TR have a known role only in cardiac or muscle development (Ybx1, Kdm6b, Zfp238, Ncoa4, Hoxa3, Camta1, Mdfi, Sin3b, Rbl2), showing their involvement in cardiac functions in the adult. In addition, several other highly expressed TR are poorly known in muscle such as the cofactors Ctbp1, Dpf2, Ewsr1 or Puf60 in mice, and Rere or Tceb2 in rats. These results suggest that they may be of importance in the normal adult heart.

In contrast, some TR++ or TR-- are involved in heart function or even regulate directly the expression of CC and/or EnProd genes. The transcription factor Epas1, also known as Hif2a, that was found in CTR rats and Isop mice, has a role in heart function and energy metabolism, controlling production of hypoxia-induced genes [28]. In both rat groups, we identified the coactivator Ankrd1, also known as

CARP, which plays an important role in cardiogenesis and also acts as stretch receptor [29]. Interestingly Ankrd1 is known to be a binding partner of Ybx1 that was identified in Ctrl mice. The transcription factor Ybx1 is involved in the transcription of myosin light-chain 2 and regulates mitochondrial oxidative phosphorylation gene expression [30]. We identified in mice only the cofactor Fhl2, which codes for a LIM-only protein and is known to interact with Ctnnb1, ß-catenin, during muscle differentiation [31]. Similarly Mdfi interacts with Ctnnb1 (protein-protein interaction database STRING, http://string-db.org/). We also detected two cofactors involved in chromatin modification that are known in heart function or muscle development: Hdac6 and Sin3b regulating histone acetylation [32,33]. In addition, Sin3b is part of a stable complex in mammals with Morf4l1, alias Mrg15, that we also found as TR++ [34]. Finally, Nrip1, detected in Ctrl mice, is involved in transcription of several EnProd genes in skeletal muscle [35], and Purb, detected in LVH rats, regulates cardiac isoforms of actin and myosin [36].

Sixteen TR were DNA-binding transcription factors with binding sites in many promoters of EnProd and CC genes, which were statistically over-represented for 6 of them (Ybx1, Eomes, Nfyc, Epas1, Myb, and Purb). Among them, those having enriched binding sites in both EnProd and CC gene promoters (Ybx1, Nfyc, Myb) may be direct determinants of EnProd and CC coupling, although the role of Nfyc and Myb is still not known in the adult heart.

EnProd and CC genes involved in coupled coordination

In each group, the EnProd and CC genes correlated with TR++ were more numerous than with TR-- but quite similar (more than 83% common). This core set of EnProd and CC genes included much higher proportion of EnProd and CC genes in mice than in rats: 24% to 72% in mice and 14 to 17% in rats, probably due to the larger size of mice groups. **Figure 5** gives the correlation networks showing the coupling of parts of EnProd and CC via TR++ or TR--. The core set of EnProd and CC genes was particularly large and stable in mice since 95% (157 genes) of the core set found in Isop mice was part of the core set in Ctrl mice. This was less clear in rats since only 45% (24 genes) of the core set in LVH rats was common with CTR rats. In addition, the core set of EnProd and CC genes was well-clustered

with a mean correlation between genes within the set between 0.35 and 0.66 across the groups, further showing the coupling of the two functions. Similarly, the sets of TR++ or TR-- were each highly clustered (mean correlation 0.53 to 0.81 for TR++, and 0.69 to 0.92 for TR--) and logically inversely correlated each other (mean inter-correlation -0.57 to -0.66).

Moreover, an emergent core set including 16 EnProd and CC genes was found in the four groups, and it could be considered as the coordination core of EnProd and CC. It contained 7 EnProd genes including Ndufs2, coding for a subunit of the mitochondrial respiratory chain complex I, Cs and Suclg1 part of the tricarboxylic acid cycle, two genes from the end-stage of glycolysis (Ldhb, Pdha1), and two genes of the first stage of gluconeogenesis (Mdh1) or fatty acid oxidation (Acadvl). This core also included 9 CC genes coding for the main contractile proteins of cardiac muscle (Actc1, Myl2, and Myl3), together with the regulatory proteins troponin (Tnnc1, Tnnt2) and tropomyosin (Tpm1), as well as genes involved in Ca^{2+} transport and storing (Cacna1c, and Atp2a2 alias Serca2) or mitochondrial ATP transport (Slc25a4 alias Ant).

TR as potential determinants of EnProd and CC inverse coupling

Table 2 gives the TR+- and TR-+ potentially reflecting the opposite coupling of EnProd and CC, which were found only in rats and were 2-fold more numerous in LVH than CTR rats. They involved two transcription factors common to the two groups, Meox2 and Elk3. Elk3 has been shown as a transcriptional repressor of Myh6 and is involved in the transcriptional response to hypoxia [37]. In LVH rats, six of these TR have already been associated to cardiac hypertrophy participating in compensatory events: E2f6, Esrra, Ccnd1, Creg1, Hopx, or Tgfb1i1 [38–43]. We also identified the transcription factors Wt1 and Rreb1, which appeared of particular interest due to their enrichment in binding sites in EnProd and CC gene promoters. Although not known in cardiac hypertrophy, their presence appears coherent according to their functional role related to the control of Ca^{2+} homeostasis or to cell adhesion [44,45].

Validation of some TRs with ChIP data

We used the ChIP Enrichment Analysis web tool (ChEA, http://amp.pharm.mssm.edu/lib/chea.jsp)

that includes databases manually created from literature or obtained after reprocessing of raw data files of ChIP experiments in human and mouse cells [46,47]. A total of 206 transcription factors were present in these databases including only 8 ones among the 24 DNA-binding TR that we analyzed with Genomatix. The databases also included the histone deacetylase Sin3b, and the cofactors Ccnd1 and Ctnnb1. For each function, we analyzed the enrichment of targets against the database. For EnProd, 10 factors were enriched in targets in the EnProd gene list: Ccnd1, Ctnnb1, Crx, Eomes, Irf1, Myb, Sin3b, Stat1, Tfap2a, and Tcf7. For the CC function, 8 factors were enriched in targets in the CC gene list: Ctnnb1, Crx, Eomes, Myb, Sin3b, Tfap2a, Tcf7 and Wt1. After filtering these results on mouse experiments, we could validate Crx, Eomes, Myb, Sin3b, and Tcf7 that we identified from mice data. Additional results were obtained from ChIP-seq experiments of the ENCODE project based on various human cell lines. Only 5 of the transcription factors that we selected (E2F6, ESRRA, IRF1, RUNX3, and STAT1) have been studied. The results show that ESRRA, IRF1, and STAT1 had enriched targets within the EnProd genes but none had enriched targets within the CC genes. Detailed results are given in Supplementary Table 4.

Discussion

In this work based on gene co-expression analysis from microarray data, we proposed a global strategy to uncover emergent properties of two major cardiac cellular functions. In all of the studied groups, each function was part of a coordination network including positive and inverse associations with external genes. We highlighted for the first time at the mRNA level a constant coupling between EnProd and CC already known at the physiological level. It could be explained by combinations of TR coordinated with both functions. Unexpectedly, those combinations were very different across the groups showing the versatility of the adaptive transcriptional control as opposed to the stability of the functional coupling. In particular, the set of TR specific of LVH rats, including several already individually known TR in cardiac hypertrophy, may be of importance in this pathology.

Coordination as an emergent property of a function

The concept of emergence was essential in our work because we gave priority to collective properties of gene sets (coordination) instead of individual gene properties (expression).

We used a two-stage statistical filter to make emerge coordination, which included 1. the selection of genes significantly correlated with a functional gene set using the Bonferroni correction for multiple testing, and 2. the selection of genes coordinated with the functional gene set based on the count of their significant correlations. In post-genomic studies, the word "coordination" is used in various very different meanings, such as simultaneous cooperative activity of several gene products in a specific process, in particular the transcription process [9,48,49], or expression differences of gene pairs over a set of multiple biological perturbations [50], or functions sharing common components with specific properties [6]. Here we proposed a new operational definition of coordination applied to co-expression of one gene with a functional gene set, which relies on the number of correlations of the gene with the gene set beyond chance.

Interpretation of correlation between transcripts

In the different groups studied here, the observed correlations between transcripts, either positive or negative, were strikingly strong and numerous. We previously showed similar numbers of positive and negative correlations in the entire gene network, which had very different functional interpretation [17]. In this work, we also studied positive and negative correlations independently, whereas most of the works dealing with co-expression focused only on positive correlations, or even dealt with the absolute value of correlation coefficient [13,14]. The frequent lack of interest for negative correlations may come from the fact that they seem hard to interpret, and that they have been described as poorly reproducible [51].

Our work resembles the clustering analyses already performed in mouse or human cardiomyopathies [11–13]. In these studies, co-expression modules were associated with the predominant function of the genes within the module. This way, authors highlighted on emergent modules related to energy metabolism, muscle contraction, as well as translation or response to cardiac stress [11]. Other authors

used the weighted co-expression network analysis to show the lack of a unique global fetal gene expression program in failing and hypertrophied mouse hearts [12]. However, such expression modules were difficult to clearly interpret at the functional level and to compare between groups. In addition, these studies did not give detailed data about coordination between several modules. This is why we chose to start our analyses from *a priori* known functional gene sets.

Coordination with EnProd and CC: functional coupling

Our results show the integration of the studied functions in a broad network. Because the genes inversely coordinated with EnProd or CC were all well-expressed, the inverse coordination could not reflect important repression but more likely fine modulation necessary for stabilizing the gene network. Indeed, from the viewpoint of systems theory, negative regulations are critical to ensure the stability of a network of positive controls [52]. For each function, the Coord and InvCoord genes largely differed between the groups but inside the function the involved genes were quite stable, reflecting a certain robustness of the functional coordination. These changes across cardiac functional states reflect the substantial reorganization of the co-expression network in response to chronic pressure overload or to acute adrenergic stimulation. This reorganization was not associated to dramatic changes in gene expression, as no expression change was detected in rats whose cardiac hypertrophy was moderate [18]. This is accordance with the work of Dewey *et al*, which showed different expression modules in fetal, normal adult, hypertrophic, and failing hearts [12]. Such results show the interest of coordination analysis with gene co-expression, which gives much more information than the study of gene expression on a one-by-one basis.

For the four studied groups and the two functions, the GO families enriched by the Coord genes were mostly related to the general activity of the cardiac tissue. However, a proportion of Coord and InvCoord genes were associated to heart development and to development of other organs or tissues. Although this may partly result from a knowledge bias due to the huge number studies on development, it shows that beyond development these genes are still expressed and functional. This was already shown for some cardiac transcription factors [53,54].

The global coupling of EnProd and CC gene sets was shown here for the first time at the mRNA level. It is known at the biochemical level between mitochondrial ATP synthesis and ATP consumption by the actin-myosin cross-bridge through Ca^{2+} as signaling molecule [55,56]. In LVH rats specifically, the GO analysis of InvCoord genes unveiled an inverse coupling between EnProd and CC, reflecting important functional changes at the transcript level that are probably related to the worsening of left ventricular function at an early stage of cardiac hypertrophy [57].

Potential determinants of EnProd and CC coupling

Although the mRNA level is not the activating signal of TR, the sets of TR identified in this work may be involved, either directly or not, in the coupled coordination of EnProd and CC. It has long been accepted that co-expressed genes are somehow functionally related, and are likely to share transcriptional regulation [13,58,59]. Co-expression and promoter analyses have been frequently brought together for determining potential transcription factors shared by a co-expression module [13,26,59]. Moreover, co-expression correlations between transcription factors and their target genes have also been shown initially in lower eukaryotes [60] and then in mammals [26]. Here we were interested in TR with or without DNA-binding properties, which were co-expressed with several genes of the EnProd and CC functions, thus giving a preliminary validation about their association with these genes.

Very few TR were common to several groups, which disclosed a large variability of the potential actors of the functional coupling across groups. This variability is paradoxical because it contrasts with a kind of functional robustness highlighted across different groups by the stability of the gene core within EnProd and CC functions. This robust gene core included sixteen well-known genes of major importance in the cardiac tissue activity. Finally, our results suggest a versatility of the transcriptional control which is in accordance with the present knowledge on the combinatorial properties of transcriptional regulation [61,62], and on the plasticity of regulatory networks well-demonstrated in yeast [63].

The bibliographical analysis of TR involved in EnProd and CC coupling unveiled the lack of knowledge about transcriptional regulation in functional adult tissues. Our approach thus appears a powerful

screening method able to improve the understanding of the transcriptional regulation with new candidates to test. In particular, the ability of E2f6, Myb, Nfyc, Rreb1, Wt1, and Ybx1, to bind the promoters of EnProd and CC genes and/or to modulate their mRNA level would need to be tested in cardiac tissues or cells. Unfortunately, most of the transcription factors showing significant over-representation of binding sites with Genomatix analysis - Elk3, Epas1, Nfyc, Purb, Rreb1, and Ybx1 - could not be found in any ChIP experiment. However, ChIP data obtained in various mouse cell types, but not in cardiac cells, partly validated the Genomatix binding sites enrichment results for Eomes and Myb in mice, and supported the potential role of Crx, Sin3b, and Tcf7.

In addition, some TR were involved in the coupled inverse coordination of EnProd and CC only in rats. They were coordinated with few genes of the functions that were different from those involved in the coupled coordination (data not shown). This coupled inverse coordination was particularly important for LVH rats showing a clear reorganization of EnProd and CC coupling in this situation of moderate cardiac hypertrophy. In addition, ENCODE data could reinforce the Genomatix binding sites enrichment of Esrra. This gene as several other TR+- or TR-+ have already been associated to cardiac hypertrophy, but have never been analyzed collectively in a single study [39–43]. The simultaneous detection of several TR of importance in cardiac hypertrophy makes one of the interests of this work.

Conclusions

The global study of gene co-expression with functional gene sets unveiled new emergent properties about coordination of cardiac cellular functions, linking molecular biology and physiology. We highlighted in various conditions the coupling of energy metabolism and cardiac contraction already known at other levels, through a diversity of TR coordinated with the two functions. These results put forward the versatile combinatorial nature of the transcriptional regulation across groups in contrast with the stability of the coupling between functions ensuring a functional robustness. In addition, this work revealed a combination of specific TR that could be of importance in cardiac

hypertrophy. This work should be pursued with other tissues and other functions towards the comprehension and the modeling of gene regulatory networks.

Acknowledgments

We thank Dr. Jacques-Yuan Li and Dr. Liliana Legedz for many helpful discussions. The Affymetrix experiments made in rats were supported by grants from “Fondation de France” and the French Ministry of Higher Education and Research.

Disclosures: none.

References

1. H. Kitano, *Nature*, 2002, **420**, 206–210.
2. A. L. Barabasi and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.
3. L. Alberghina, T. Höfer, and M. Vanoni, *J. Biotechnol.*, 2009, **144**, 224–233.
4. C. van Waveren and C. T. Moraes, *BMC Genomics*, 2008, **9**, 18.
5. J. M. Baughman, R. Nilsson, V. M. Gohil, D. H. Arlow, Z. Gauhar, and V. K. Mootha, *PLoS Genet*, 2009, **5**, e1000590.
6. W. Ma, D. Yang, Y. Gu, X. Guo, W. Zhao, and Z. Guo, *Genomics*, 2009, **94**, 94–100.
7. P. K. Shreenivasaiah, S. H. Rho, T. Kim, and D. H. Kim, *J. Mol. Cell. Cardiol.*, 2008, **44**, 460–469.
8. D. Diez, A. M. Wheelock, S. Goto, J. Z. Haeggstrom, G. Paulsson-Berne, G. K. Hansson, U. Hedin, A. Gabrielsen, and C. E. Wheelock, *Mol. Biosyst.*, 2010, **6**, 289–304.
9. A. J. Lusis and J. N. Weiss, *Circulation*, 2010, **121**, 157–170.
10. S. Y. Chan and J. Loscalzo, *Circ. Res.*, 2012, **111**, 359–374.
11. U. D. Akavia and D. Benayahu, *Physiol. Genomics*, 2008, **35**, 305–315.
12. F. E. Dewey, M. V. Perez, M. T. Wheeler, C. Watt, J. Spin, P. Langfelder, S. Horvath, S. Hannenhalli, T. P. Cappola, and E. A. Ashley, *Circ. Cardiovasc. Genet.*, 2011, **4**, 26–35.
13. C. Danko and A. Pertsov, *BMC Med. Genomics*, 2009, **2**, 31.
14. B. Zhang and S. Horvath, *Stat. Appl. Genet. Mol. Biol.*, 2005, **4**, 17.
15. F. Azuaje, L. Zhang, C. Jeanty, S.-L. Puhl, S. Rodius, and D. R. Wagner, *BMC Med. Genomics*, 2013, **6**, 13.

16. M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, *Nat. Protoc.*, 2007, **2**, 2366–2382.
17. M. P. Gustin, C. Z. Paultre, J. Randon, G. Bricca, and C. Cerutti, *Physiol. Genomics*, 2008, **34**, 34–41.
18. C. Cerutti, M. Kurdi, G. Bricca, W. Hodroj, C. Paultre, J. Randon, and M. P. Gustin, *Physiol. Genomics*, 2006, **27**, 295–308.
19. J. Wong, C. Chang, R. Agrawal, G. B. Walton, C. Chen, A. Murthy, and A. J. Patterson, *Crit Care Med*, 2010, **38**, 25–31.
20. J. Sassard, M. Vincent, V. Oréa, P. Privat, and A. Bataillard, *Clin. Exp. Hypertens. N. Y. N 1993*, 1997, **19**, 567–575.
21. S. A. Doggrell and L. Brown, *Cardiovasc. Res.*, 1998, **39**, 89–105.
22. R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, *Biostat. Oxf. Engl.*, 2003, **4**, 249–264.
23. V. G. Tusher, R. Tibshirani, and G. Chu, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5116–5121.
24. B. Efron, *The jackknife, the bootstrap, and other resampling plans*, Society for Industrial and Applied Mathematics;, 1982.
25. A. Alexa, J. Rahnenführer, and T. Lengauer, *Bioinformatics*, 2006, **22**, 1600–1607.
26. Q. Gu, S. H. Nagaraj, N. J. Hudson, B. P. Dalrymple, and A. Reverter, *BMC Genomics*, 2011, **12**, 23.
27. K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, and T. Werner, *Bioinformatics*, 2005, **21**, 2933–2942.
28. J. J. Ho, J. L. Metcalf, M. S. Yan, P. J. Turgeon, J. J. Wang, M. Chalsev, T. N. Petruzziello-Pellegrini, A. K. Tsui, J. Z. He, H. Dhamko, H. S. Man, G. B. Robb, B. T. Teh, M. Ohh, and P. A. Marsden, *J Biol Chem*, 2012, **287**, 29003–29020.
29. S. Kojic, D. Radojkovic, and G. Faulkner, *Crit. Rev. Clin. Lab. Sci.*, 2011, **48**, 269–294.
30. S. Matsumoto, T. Uchiumi, H. Tanamachi, T. Saito, M. Yagi, S. Takazaki, T. Kanki, and D. Kang, *Biochem J*, 2012, **443**, 573–584.
31. B. Martin, R. Schneider, S. Janetzky, Z. Waibler, P. Pandur, M. Kühl, J. Behrens, K. von der Mark, A. Starzinski-Powitz, and V. Wixler, *J. Cell Biol.*, 2002, **159**, 113–122.
32. van Oevelen, C., C. Bowman, J. Pellegrino, P. Asp, J. Cheng, F. Parisi, M. Micsinai, Y. Kluger, A. Chu, A. Blais, G. David, and B. D. Dynlacht, *Mol Cell Biol*, 2010, **30**, 5686–5697.
33. D. D. Lemon, T. R. Horn, M. A. Cavasin, M. Y. Jeong, K. W. Haubold, C. S. Long, D. C. Irwin, S. A. McCune, E. Chung, L. A. Leinwand, and T. A. McKinsey, *J. Mol. Cell. Cardiol.*, 2011, **51**, 41–50.
34. P. Jelinic, J. Pellegrino, and G. David, *Mol Cell Biol*, 2011, **31**, 54–62.

35. B. C. Frier, C. R. Hancock, J. P. Little, N. Fillmore, T. A. Bliss, D. M. Thomson, Z. Wan, and D. C. Wright, *J. Appl. Physiol.*, 2011, **111**, 688–695.
36. J. Ji, G. L. Tsika, H. Rindt, K. L. Schreiber, J. J. McCarthy, R. J. Kelm, and R. Tsika, *Mol. Cell. Biol.*, 2007, **27**, 1531–1543.
37. T. Serchov, H. Dubois-Pot-Schneider, C. Charlot, F. Rösl, and B. Wasylyk, *J. Biol. Chem.*, 2010, **285**, 21223–21232.
38. L. Xu, J. M. Liu, and L. Y. Chen, *J. Hypertens.*, 2004, **22**, 1579–1587.
39. J. M. Huss, K. ichi Imahashi, C. R. Dufour, C. J. Weinheimer, M. Courtois, A. Kovacs, V. GiguŠre, E. Murphy, and D. P. Kelly, *Cell Metab.*, 2007, **6**, 25–37.
40. L. K. Mullany, P. White, E. A. Hanse, C. J. Nelsen, M. M. Goggin, J. E. Mullany, C. K. Anttila, L. E. Greenbaum, K. H. Kaestner, and J. H. Albrecht, *Cell Cycle Georget. Tex*, 2008, **7**, 2215–2224.
41. Z. Bian, J. Cai, D. Shen, L. Chen, L. Yan, Q. Tang, and H. Li, *J. Cell. Mol. Med.*, 2009, **13**, 1302–1313.
42. E. E. Yund, J. A. Hill, and R. S. Keller, *J. Mol. Cell. Cardiol.*, 2009, **47**, 520–527.
43. B. Westendorp, J. L. Major, M. Nader, M. Salih, F. H. H. Leenen, and B. S. Tuana, *FASEB J.*, 2012, **26**, 2569–2579.
44. M. Melani, K. J. Simpson, J. S. Brugge, and D. Montell, *Curr. Biol.*, 2008, **18**, 532–537.
45. M. F. Ritchie, Y. Zhou, and J. Soboloff, *Cell Calcium*, 2011, **49**, 314–321.
46. A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, and A. Ma'ayan, *Bioinformatics*, 2010, **26**, 2438–2444.
47. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan, *BMC Bioinformatics*, 2013, **14**, 128.
48. S. Komili and P. A. Silver, *Nat. Rev. Genet.*, 2008, **9**, 38–48.
49. A. A. Petti and G. M. Church, *Genome Res.*, 2005, **15**, 1298–1306.
50. H. Less and G. Galili, *BMC Syst. Biol.*, 2009, **3**, 14.
51. H. K. Lee, A. K. Hsu, J. Sajdak, J. Qin, and P. Pavlidis, *Genome Res.*, 2004, **14**, 1085–1094.
52. L. von Bertalanffy, *General System Theory: Foundations, Development, Applications*, George Braziller Inc., New York, 1968.
53. B. D. Thattaliyath, C. B. Livi, M. E. Steinhelper, G. M. Toney, and A. B. Firulli, *Biochem. Biophys. Res. Commun.*, 2002, **297**, 870–875.
54. S. Pikkarainen, H. Tokola, R. Kerkela, and H. Ruskoaho, *Cardiovasc. Res.*, 2004, **63**, 196–207.
55. Y. Yaniv, W. C. Stanley, G. M. Saidel, M. E. Cabrera, and A. Landesberg, *Ann N Acad Sci*, 2008, **1123**, 69–78.
56. R. S. Balaban, *Biochim. Biophys. BBA - Bioenerg.*, 2009, **1787**, 1334–1341.
57. D. L. Mann, *Circulation*, 2005, **111**, 2837–2849.

58. M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, *Proc. Natl. Acad. Sci.*, 1998, **95**, 14863–14868.
59. E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman, *Nat Genet*, 2003, **34**, 166–176.
60. T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood, *Science*, 2001, **292**, 929–934.
61. T. Ravasi, H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C. O. Daub, A. R. R. Forrest, J. Gough, S. Grimmond, J.-H. Han, T. Hashimoto, W. Hide, O. Hofmann, A. Kamburov, M. Kaur, H. Kawaji, A. Kubosaki, T. Lassmann, E. van Nimwegen, C. R. MacPherson, C. Ogawa, A. Radovanovic, A. Schwartz, R. D. Teasdale, J. Tegnér, B. Lenhard, S. A. Teichmann, T. Arakawa, N. Ninomiya, K. Murakami, M. Tagami, S. Fukuda, K. Imamura, C. Kai, R. Ishihara, Y. Kitazume, J. Kawai, D. A. Hume, T. Ideker, and Y. Hayashizaki, *Cell*, 2010, **140**, 744–752.
62. F. Spitz and E. E. M. Furlong, *Nat. Rev. Genet.*, 2012, **13**, 613–626.
63. H. Li and A. D. Johnson, *Curr. Biol.*, 2010, **20**, R746–R753.

Figure legends

Figure 1. Overview of the analysis of functional coordination. **A.** Flow diagram describing the sequence of statistical and biological knowledge filters used in the analysis. **B.** For a given functional gene set F, the genes of the microarray positively or negatively correlated with more genes of F than randomly are called coordinated (Coord) or inversely coordinated (InvCoord) with F. Subsets of Coord and InvCoord genes delineated by small ellipses are involved in known cellular functions, illustrating the coordination of F with several other functions. Arrows denote coordination corresponding to multiple correlations with F genes. **C.** Considering two functional gene sets F_1 and F_2, genes known as transcriptional regulators (TR) are identified among the coordinated genes common to F_1 and F_2 with similar (TR++ and TR--) or opposite signs (TR+- and TR-+). Due to the mathematical transitivity of the correlation, TR++ and TR-- are associated with the same F_1 and F_2 genes (large rectangles within F_1 and F_2) reflecting coupled coordination of F_1 and F_2. Differently, TR+- and TR-+ are associated with other F_1 and F_2 genes (small rectangles within F1 and F2), reflecting coupled inverse coordination of F_1 and F_2. Oriented arrows show the potential causality, either direct or indirect, between TR and the functional gene sets.

Figure 2. Histogram curve of the number of positive correlations of each microarray gene with the energy production gene set or with a random gene set of equal size obtained in control rats. The histogram is represented for the genes correlated with at least one gene of the function. Correlations were computed here without any resampling procedure and significance was taken at $p<10^{-4}$ ($r > 0.8201$ for $n=16$). Less than 3 correlations with the gene set was not significantly (NS) different from what was obtained with a random gene set. Significant differences vs random were obtained for numbers of correlations greater than 3.

Figure 3. Coordination of the cardiac contraction gene set in control rats. The figure shows the correlation links of the 18 coordinated (Coord) genes positively correlated with at least 9 genes of the

function (red edges), and of the 16 inversely coordinated (InvCoord) genes correlated with at least 4 genes of the function (green edges). Twenty-two genes of the function are involved in these links, some of them being similarly involved in the coordination and the inverse coordination.

Figure 4. Families of Gene Ontology (GO) terms of the Biological Process category enriched in coordinated genes (Coord) for the two studied functions, cardiac contraction and energy production, in rats and mice. For each group and each function, the number of genes of the function is indicated in rectangles linked (+) to ellipses containing the number of Coord genes. The enriched GO terms were manually grouped together according to 4 families: energy metabolism, muscle processes, protein processes, and non-muscle development. The percentage of Coord genes that enriched GO terms from one family are represented as follows: □ no gene, ■0-10%, ■ 10-20%, ■ > 20%.

Figure 5. Coordination networks showing the coupling of energy production and cardiac contraction functions via common transcriptional regulators positively and inversely coordinated with both functions. **A**. control rats; **B**. hypertensive rats; **C.** control mice; **D.** isoprenaline-infused mice.
The red edges denote the positive correlations and the green edges denote the negative ones. Elliptic nodes refer to the genes of the functions, and diamond nodes refer to transcriptional regulators. Yellow nodes are those present in control and pathological groups, and orange nodes are those common to the four groups. The networks were obtained with the free software Cytoscape 2.8.1 (http://www.cytoscape.org). EnProd: energy production; CC: cardiac contraction; TR++: transcriptional regulators positively coordinated with the two functions; TR--: transcriptional regulators inversely coordinated with the two functions.

Table 1: Transcriptional regulators involved in the coupled coordination of energy production and cardiac contraction in rats and mice.

Group	Transcriptional regulator			% genes with binding sites		Functional role						Reference
						Heart or muscle		Energy	General development	Inflammation	Other	
	type	symbol	expr (centile)	EnProd	CC	function	development					
Ctrl mice	+ +	Ctnnb1	100			■	■					Rao TP, 2010
Ctrl mice	+ +	Fhl2	100			■	■					Martin B, 2002 [31]
Ctrl mice	+ +	Hopx	99			■	■					Trivedi CM, 2011
Ctrl mice	+ +	Rac1	99			■						Elnakish MT, 2011
Ctrl mice	+ +	*Hdac6*	78			■						Lemon DD, 2011 [33]
Ctrl mice	+ +	**Ybx1**	100	31 *	34 *		■	■				Matsumoto S, 2012 [30]
Ctrl mice	- -	**Hoxa3**	47	22	9		■					Diman NY, 2011
Ctrl mice	+ +	Camta1	95				■					Muller-Borer B, 2012
Ctrl mice	- -	Mdfi	51				■					Pan W, 2006
Ctrl mice	+ +	*Sin3b*	91				■					Jelinic P, 2011 [34]
Ctrl mice	+ +	Nrip1	86					■				Frier BC, 2011 [35]
Ctrl mice	- -	**Eomes**	47	8	25 *				■			Teo AK, 2011
Ctrl mice	- -	**Runx3**	46	28	26				■			Fu Y, 2011
Ctrl mice	- -	**Tcfap2a**	51	9	17				■			Bragança J, 2003
Ctrl mice	+ +	Lrrfip1	87							■		Lee YH, 2006
Ctrl mice	+ +	**Nfyc**	93	41 *	41 *						■	Murai-Takeda A, 2010
Ctrl mice	- -	**Crx**	46	23	21						■	Hao H, 2012
Ctrl mice	- -	Batf2	60								■	Su ZZ, 2008
Ctrl mice	+ +	Crebzf	82								■	Xie YB, 2009
Ctrl mice	+ +	Ctbp1	97								■	Vernochet C, 2009
Ctrl mice	+ +	Dpf2	90								■	Matsuyama R, 2010
Ctrl mice	+ +	Ewsr1	92								■	Leeman-Zakaryan RP, 2009
Ctrl mice	- -	Med18	88								■	Mukundan B, 2011
Ctrl mice	+ +	Puf60	94								■	Hastings ML, 2007
Ctrl mice	+ +	*Morf4l1*	100								■	Xie T, 2012
Ctrl mice	- -	Nrip2	71									-
Isop mice	+ +	**Epas1**	99	31 *	20	■		■				Ho JJ, 2012 [28]
Isop mice	+ +	Fhl2	100			■	■					Martin B, 2002 [31]
Isop mice	- -	**Tcf7**	58	21	22				■			Mao CD, 2011
Isop mice	- -	**Myb**	44	40 *	39 *						■	Kolodziejska KM, 2008
CTR rats	+ +	**Epas1**	84	31 *	20	■		■				Ho JJ, 2012 [28]
CTR rats	+ +	Ankrd1	100			■	■					Kojic S, 2011 [29]
CTR rats	+ +	**Stat1**	86	30	27	■				■		Mir SA, 2012
CTR rats	+ +	**Zfp238**	57	11	12		■					Yokoyama S, 2011
CTR rats	- -	Ncoa4	92				■					Kollara A, 2012
CTR rats	+ +	**Nfe2l1**	61	11	11			■	■			Koch A, 2011
CTR rats	+ +	**Thrb**	40	4	9			■				Arsanjani R, 2011
CTR rats	+ +	*Kdm6b*	61						■	■		Jiang W, 2012

CTR rats	+ +	Rere	88			Wang L, 2008
CTR rats	- -	Tceb2	97			Hwang J, 2011
LVH rats	+ +	Ankrd1	100			Kojic S, 2011
LVH rats	+ +	**Purb**	74	15	26 *	Ji J, 2007 [36]
LVH rats	+ +	Rbl2	65			Sdek P, 2011
LVH rats	- -	Men1	60			Zhang HL, 2007
LVH rats	+ +	**Phtf2**	84	no matrix		Manuel A, 2000

EnProd: energy production; CC: cardiac contraction; CTR: control rats; LVH: rats with left ventricular hypertrophy; Ctrl: control mice; Isop: isoproterenol-treated mice.
Gene symbols of DNA-binding transcription factors are in bold, and those of chromatin modifiers are in italic.
Percentages of binding sites with * indicate significant over-representation of binding sites ($p<0.05$); "no matrix" means that no individual position weight matrix was defined for the transcription factor.
The role of each transcriptional regulator in heart or muscle function or development, energy metabolism, inflammation, general development or in other function excluding cancer is indicated in grey. For each transcriptional regulator, the most recent reference is given (for those not linked to a reference number, i.e. not cited in the text, see Supplementary Information).

Table 2: Transcriptional regulators involved in the coupled inverse coordination of energy production and cardiac contraction in rats.

Group	Transcriptional regulator			% genes with binding sites		Functional role						Reference
	type	symbol	expr (centile)	EnProd	CC	Heart or muscle function	Heart or muscle development	Energy	General development	Inflammation	Other	
CTR rats	- +	**Elk3**	83	49 *	26							Serchov T, 2010 [37]
CTR rats	+ -	**Dbp**	79	23	14							Wang Q, 2010
CTR rats	- +	Mdfic	58									Reiss-Sklan E, 2009
CTR rats	- +	**Irf1**	75	17	25							Saha B, 2010
CTR rats	- +	**Litaf**	82	no matrix								Tang X 2011
CTR rats	- +	**Meox2**	66	no matrix								Douville JM 2011
LVH rats	- +	**Elk3**	83	49 *	26							Serchov T, 2010 [37]
LVH rats	- +	**Meox2**	68	no matrix								Douville JM; 2011
LVH rats	- +	Hopx	94									Trivedi CM, 2011
LVH rats	- +	Tgfb1i1	81									Yund EE, 2009 [42]
LVH rats	+ -	Creg1	80									Bian Z, 2009 [41]
LVH rats	+ -	**E2f6**	95	39 *	33 *							Westendorp B, 2012 [43]
LVH rats	+ -	**Esrra**	76	34 *	19							Huss JM, 2007 [39]
LVH rats	- +	Ccnd1	90									Mullany LK, 2008 [40]
LVH rats	- +	**Wt1**	58	40 *	56 *							Ritchie MF, 2011 [45]
LVH rats	- +	*Hdac7*	66									Zhou B, 2011
LVH rats	- +	Phc2	84									Wu H, 2010
LVH rats	+ -	**Rreb1**	71	45 *	56 *							Melani M, 2008 [44]

EnProd: energy production; CC: cardiac contraction; CTR: control rats; LVH: rats with left ventricular hypertrophy. Gene symbols of DNA-binding transcription factors are in bold, and those of chromatin modifiers are in italic. Percentages of binding sites with * indicate significant over-representation of binding sites ($p<0.05$); "no matrix" means that no individual position weight matrix was defined for the transcription factor.

The role of each transcriptional regulator in heart or muscle function or development, energy metabolism, inflammation, general development or in other function excluding cancer is indicated in grey. For each transcriptional regulator, the most recent reference is given (for those not linked to a reference number, i.e. not cited in the text, see Supplementary Information).

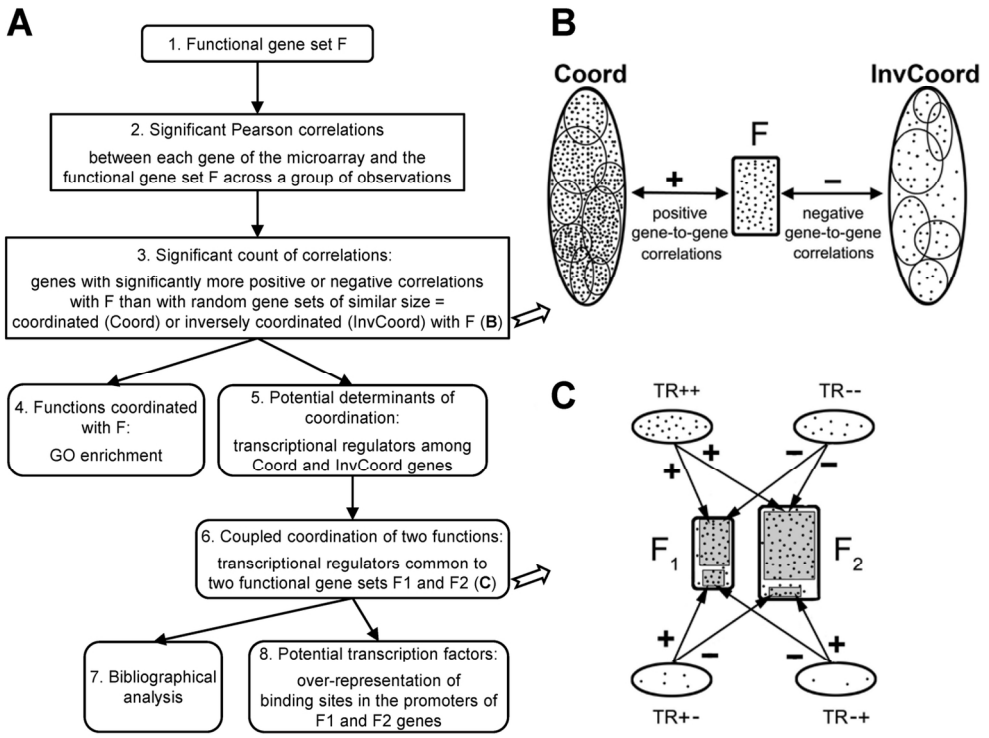


Figure 1. Overview of the analysis of functional coordination. A. Flow diagram describing the sequence of statistical and biological knowledge filters used in the analysis. B. For a given functional gene set F, the genes of the microarray positively or negatively correlated with more genes of F than randomly are called coordinated (Coord) or inversely coordinated (InvCoord) with F. Subsets of Coord and InvCoord genes delineated by small ellipses are involved in known cellular functions, illustrating the coordination of F with several other functions. Arrows denote coordination corresponding to multiple correlations with F genes. C. Considering two functional gene sets F1 and F2, genes known as transcriptional regulators (TR) are identified among the coordinated genes common to F1 and F2 with similar (TR++ and TR--) or opposite signs (TR+- and TR-+). Due to the mathematical transitivity of the correlation, TR++ and TR-- are associated with the same F1 and F2 genes (large rectangles within F1 and F2) reflecting coupled coordination of F1 and F2. Differently, TR+- and TR-+ are associated with other F1 and F2 genes (small rectangles within F1 and F2), reflecting coupled inverse coordination of F1 and F2. Oriented arrows show the potential causality, either direct or indirect, between TR and the functional gene sets.
129x99mm (300 x 300 DPI)

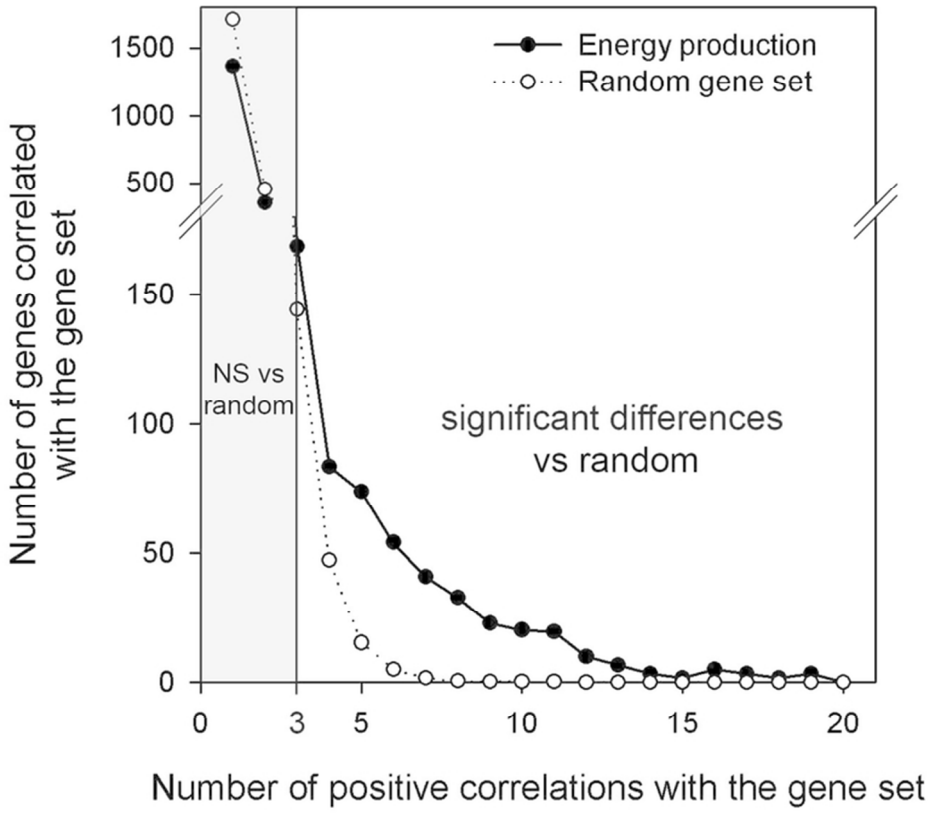


Figure 2. Histogram curve of the number of positive correlations of each microarray gene with the energy production gene set or with a random gene set of equal size obtained in control rats. The histogram is represented for the genes correlated with at least one gene of the function. Correlations were computed here without any resampling procedure and significance was taken at $p<10\text{-}4$ ($r > 0.8201$ for $n=16$). Less than 3 correlations with the gene set was not significantly (NS) different from what was obtained with a random gene set. Significant differences vs random were obtained for numbers of correlations greater than 3.
70x59mm (300 x 300 DPI)

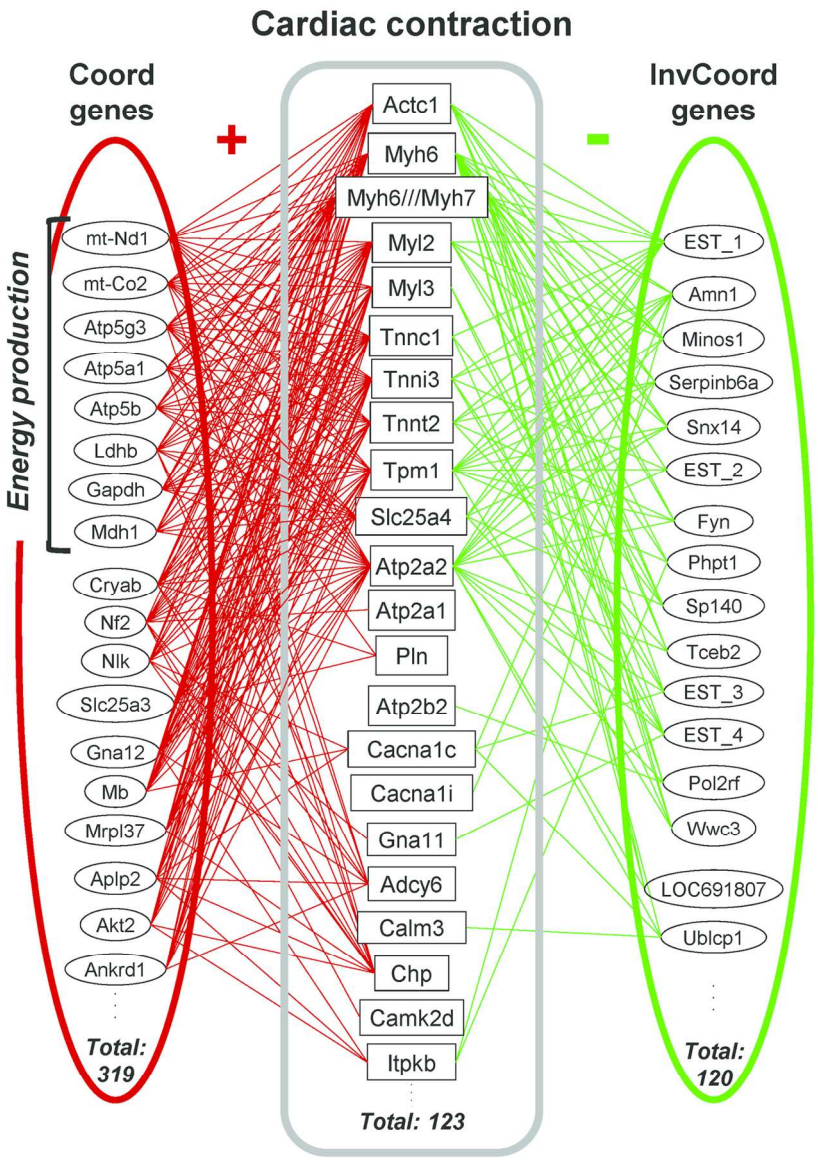


Figure 3. Coordination of the cardiac contraction gene set in control rats. The figure shows the correlation links of the 18 coordinated (Coord) genes positively correlated with at least 9 genes of the function (red edges), and of the 16 inversely coordinated (InvCoord) genes correlated with at least 4 genes of the function (green edges). Twenty-two genes of the function are involved in these links, some of them being similarly involved in the coordination and the inverse coordination.
119x173mm (300 x 300 DPI)

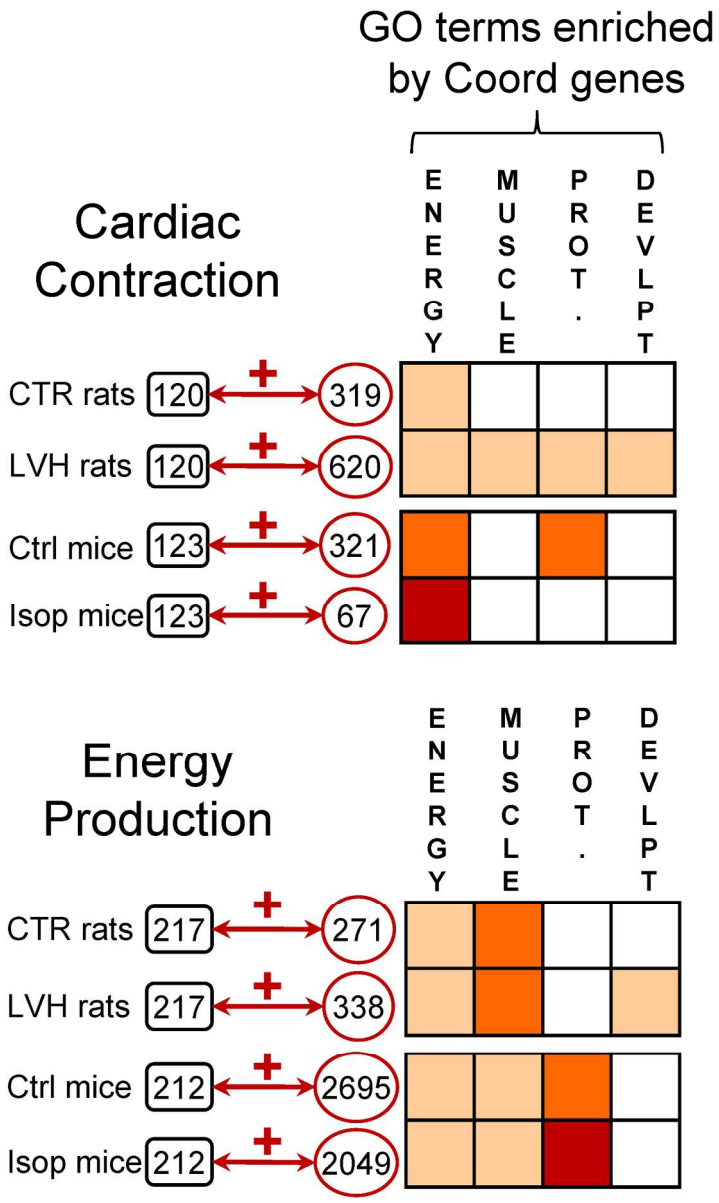


Figure 4. Families of Gene Ontology (GO) terms of the Biological Process category enriched in coordinated genes (Coord) for the two studied functions, cardiac contraction and energy production, in rats and mice. For each group and each function, the number of genes of the function is indicated in rectangles linked (+) to ellipses containing the number of Coord genes. The enriched GO terms were manually grouped together according to 4 families: energy metabolism, muscle processes, protein processes, and non-muscle development. The percentage of Coord genes that enriched GO terms from one family are represented as follows: ≤ no gene, ’0-10%, ’ 10-20%, ’ > 20%.
139x236mm (300 x 300 DPI)

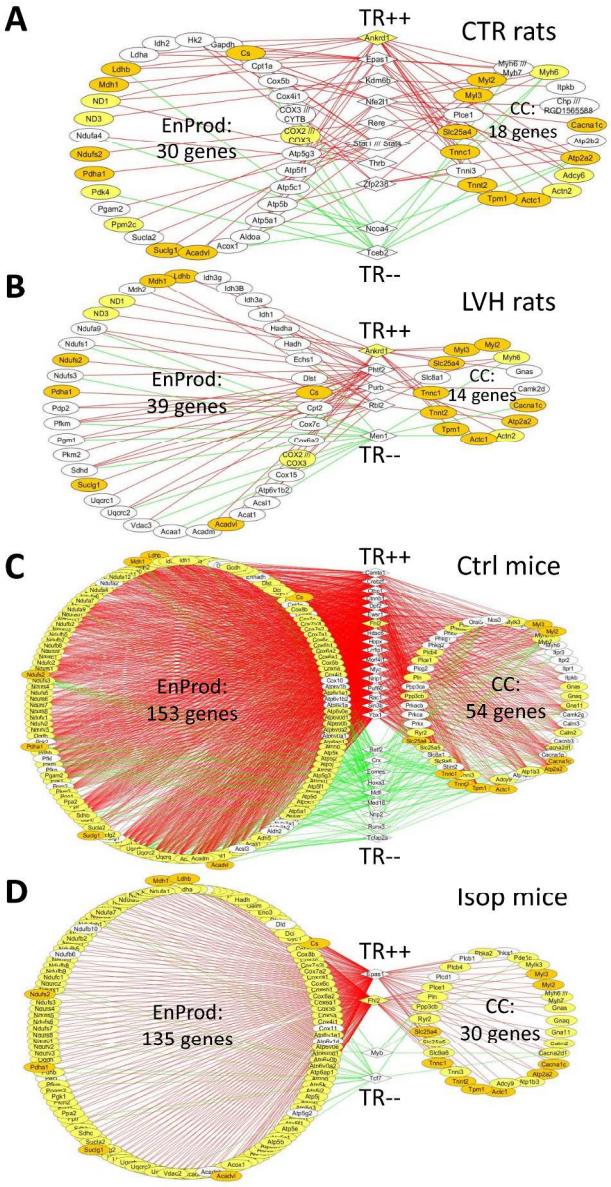


Figure 5. Coordination networks showing the coupling of energy production and cardiac contraction functions via common transcriptional regulators positively and inversely coordinated with both functions. A. control rats; B. hypertensive rats; C. control mice; D. isoprenaline-infused mice.
The red edges denote the positive correlations and the green edges denote the negative ones. Elliptic nodes refer to the genes of the functions, and diamond nodes refer to transcriptional regulators. Yellow nodes are those present in control and pathological groups, and orange nodes are those common to the four groups. The networks were obtained with the free software Cytoscape 2.8.1 (http://www.cytoscape.org). EnProd: energy production; CC: cardiac contraction; TR++: transcriptional regulators positively coordinated with the two functions; TR--: transcriptional regulators inversely coordinated with the two functions.
229x423mm (300 x 300 DPI)