

A systematic review of large scale and heterogeneous gene array data in heart failure

Citation for published version (APA):

Sharma, U. C., Pokharel, S., Evelo, C. T., & Maessen, J. G. (2005). A systematic review of large scale and heterogeneous gene array data in heart failure. *Journal of Molecular and Cellular Cardiology*, 38(3), 425-432. <https://doi.org/10.1016/j.yjmcc.2004.12.016>

Document status and date:

Published: 01/01/2005

DOI:

[10.1016/j.yjmcc.2004.12.016](https://doi.org/10.1016/j.yjmcc.2004.12.016)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Review article

A systematic review of large scale and heterogeneous gene array data in heart failure

Umesh C. Sharma^{a,b,*}, Saraswati Pokharel^b, Chris T.A. Evelo^c, Jos G. Maessen^a

^a Pump Failure Investigation Group, Department of Cardio-thoracic Surgery, Academic Hospital Maastricht, Maastricht, The Netherlands

^b Hypertension and Vascular Research Division, Henry Ford Hospital, 2799 West Grand Boulevard, Detroit, MI 48202, USA

^c Bio-informatics Division, University of Maastricht, Maastricht, The Netherlands

Received 2 October 2004; received in revised form 3 December 2004; accepted 17 December 2004

Abstract

Microarray analysis has become a widely available tool for the generation of gene expression data on a genomic scale. Since the studies with similar protocols are growing, it has become necessary to systematically revise the large body of literature to decipher the gene expression data. In this review, we analyzed and critically discussed the database presented from 14 published studies that showed the gene expression profile in heart failure (HF) using microarray as a primary tool. After comparing the diverse database from these studies, we explain the protein translational, matri-cellular, immunological and fibrosis-related mechanisms in HF. In addition to previously annotated genes, we analyzed two differentially expressed expressed sequence tags (ESTs) (KIAA0152 and Suppressor of G_{TWO} allele of the suppressor of kinetochore protein-1, SGT1) in HF and showed how bio-informatic analysis of ESTs can lead to the identification of novel pathways active in HF. We have also discussed the new publicly accessible tools that link the gene expression data to gene ontology (GO) and functionality. Finally, we have systematically revised the chromosomal localization of the genes that are specifically up-regulated in HF. We have thus spotted chromosome 1, 2, 11 and 12 as the chromosomal hotspots of HF. This methodical approach will simplify the existing concepts on the evolution and progression of HF and lead us toward the development of newer diagnostic and therapeutic tools. Although modeled to HF, this approach should be of broader scientific interest to elaborate multiple genes and complex pathways.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Microarray; Heart failure; Bioinformatics; Expressed sequence tags

1. Introduction

The evolution of new methods that enable large-scale expression analyses, like serial analysis of gene expression (SAGE), in silico analysis of expressed sequence tag (EST) databases, cDNA and oligonucleotide microarrays allows researchers to investigate organ- or pathology-specific transcriptional profiles [1,2]. Understanding the genetic basis of complex cardiovascular diseases has been a challenging task for many years and this notion of complexity is being gradually expounded with the advent and use of microarray technology in human cardiovascular diseases [3,4].

Although, study of individual signaling pathways at the genetic level has provided significant insight into the devel-

opment of heart failure (HF), approaches such as gene knock-out are limited by the interdependence of the cellular systems. A closer look on the functional genomics data shows a dynamic interaction within and across several cellular signaling and metabolic pathways. Therefore, with the exception of hypothesis-driven efforts to detect predictors, markers and modifiers of HF, the expression data should be more meaningful in the context of a detailed description of the conditions under which they are generated, with regard to the particular etiology and state of the failing hearts. Finally, the reported microarray data in HF are on varied platforms and experimental designs, thus making an integration of these data an error-prone exercise.

We believe that it is necessary to develop a new strategy to synthesize the microarray results obtained from diverse cases and models. Different arrays need to be systematically compared and most consistent observations defining the same or

* Corresponding author. Tel. +1 313 916 9141; fax: +1 313 916 1479.

E-mail address: usharma1@hfhs.org (U.C. Sharma).

similar biological phenomenon need to be translated to putative mechanisms contributory to the disease pathogenesis. Using the evidence gathered from such a cathartic approach, we reveal how the conventional method of reporting the gene array data are often ‘stripped’ of an adequate interpretation and how the information, which otherwise could be crucial to explain the mechanisms contributory to HF, is not discussed. To this end, we have revised 14 recent study reports pertaining to gene array analysis in HF [5–18]. ESTs constitute the essential tools for rapid gene discovery but are often omitted from the primary analyses. Therefore, we chose two differentially expressed ESTs in HF as model nucleotide probes in order to predict their molecular architecture and function [8,19].

Gene expression data can be analyzed on at least three levels of increasing complexity [20]. First, the level of single genes, where one seeks to establish whether each gene in isolation behaves differently in a control versus a treatment situation. The second level considers gene combinations, where clusters of genes are analyzed in terms of common functionalities, interactions, co-regulations and so forth. The third level attempts to infer the underlying regulatory regions and gene/protein networks that ultimately are responsible for the patterns observed [21–23]. Inherent to the large database and

diverse nature of the expression profile, scientists often resort to an in silico approach for the data analysis. No matter at what level the information is drawn out of the gene chips, the most important step is adequate data interpretation where the précis made from all levels of data analysis is re-evaluated, summarized and integrated to biological pathways immediately relevant to a particular disease process.

2. Heterogeneity of the study selection criteria and array design

Although the microarray studies we revised are highly heterogeneous in terms of study subjects, etiology and stage of HF and data analysis and validation procedure, the investigators centered to a common research question, i.e. the profile of differentially expressed genes in failing compared to functionally preserved hearts (Table 1). The spotlight of the researcher mainly dictates the technological description and fulfillment of the pre-requisites of a microarray experiment. It is unlikely; however, that such a system would be effective or scalable. It is recommended that the microarray data need to be standardized based on minimum information about a microarray experiment (MIAME) criteria developed by the

Table 1

Characteristics, design and results of the selected high-density arrays performed within last 5 years to discern the HF-related genes

Year	Author	Array Design	Gene Description
2000	Yang et al.	Affymetrix Hu6800 GeneChip, 7085 genes, ICM and DCM pts	Gelsolin, Elongation factor-2
2001	Barrans et al.	In-house CardioChip, 10,368 genes, HCM Pts	Elongation factor-1, CTGF, Gelsolin
2001	Jin et al.	Affymetrix Rat Genome Arrays, MI in rat models	ANP, BNP, Collagen I, Fibronectin, Biglycan, Thrombospondin-4, Lysyl oxidase, complement B/C4
2002	Barrans et al.	In-house CardioChip, 10,848 genes, DCM pts	ANP, EST (T60005), Elongation factor-1, Collagen α 1, Fibronectin, Natural killer cell enhancing factor
2002	Hwang et al.	In-house, 10,272 genes, DCM and HCM pts	DCM: ANP, Collagen α 1, Decorin, Lumican, TCR- β chain, Elongation factor-2 HCM: ANP, Decorin, Desmin, Elongation factor-2
2002	Tan et al.	Affymetrix Hu6800Fl GeneChip, 6606 genes, DCM pts	ANF, BNP, α -collagen type I, Lumican, Thrombospondin-4, CTGF
2003	Blaxall et al.	Affymetrix HuGeneChip, 6800 genes, Pre-and post LVAD pts	BNP, Collagen*, IL-8, Matrix metalloproteinase*, Tumor necrosis factor- α
2003	Boheler et al.	Incyte (Human UniGem), 10,176 genes, DCM pts	Absolute increase: Lumican, MAP2K3 Sex-dependent alteration: Calponin, MAP4K5
2003	Chen et al.	Agilent Human 1 Catalogue Array, 12,814 genes, pre- and post LVAD	Natriuretic peptide precursor B, Cathepsin B, MAPK4
2003	Grzeskowiak et al.	Human UniGene database, 30,336 genes, DCM pts	MHC-IA, Complement-I
2003	Steenman et al.	Affymetrix HG-U95A array, 12,626 genes, DCM and ICM pts	ANF, BNP, CTGF, Calponin, Biglycan, Immunoglobulin λ heavy chain
2004	Yung et al.	Affymetrix GeneChip HG-U133A, 22,283 genes, DCM pts	Collagen type XXI, Sarcoglycan, Cyclin G2
2004	Schroen et al.	Incyte Rat GEM2/3 cDNA library, 12,336 genes, Ren-2 rat model	Thrombospondin-1, Thrombospondin-2, Thrombospondin-3, Galectin-3, Collagen α 1, Osteoactivin, Osteopontin, Fibronectin, Elongation factor-1, Lumican
2004	Kittleson et al.	Affymetrix U133A microarrays, ICM pts	Immunoglobulin heavy chain, MAPK-7, Mannose receptor

ICM, ischemic cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; ANP, atrial natriuretic peptide; ANF, atrial natriuretic factor; BNP, brain natriuretic peptide; CTGF, connective tissue growth factor; TCR, T-cell receptor; IL, interleukin; MAP2K3, mitogen-activated protein kinase 3; MAP4K5, mitogen-activated protein 4 kinase 5; MHC, major histocompatibility complex.

Microarray Gene Expression Database Group (MGED; <http://www.mged.org>), a grass-root to develop standards for microarray data [24,25]. In this review, it was still difficult to grade the evidence based on MIAME criteria because of the following reasons-1. The six components of MIAME including, experimental design, array design, samples, hybridizations, measurements and normalization controls were not designed for an objective scoring, but designated as an informal specification on which microarray experiment-annotation tools may be based. 2. Some components of MIAME did not reach consensus among the MGED working group. Therefore, the peer-reviewed publications together with the information that was necessary to verify and reproduce the experimental results were considered to be eligible to be included in our data discussion.

3. Gene expression fingerprint of extracellular matrix: a common lesson from diverse arrays

In this section, we discuss how the different studies can provide a common and more convincing evidence to explain the matrix changes undergoing in failing hearts. A robust increase in the expression of ECM-related genes in most of the microarray analyses underscored the importance of myocardial matrix remodeling in failing hearts [26,27]. These genes included connective tissue growth factor (CTGF), both fibrillar and reticular collagens, along with the genes involved in controlling the deposition and orientation of collagens, and binding of cells to ECM. Recently, small leucine-rich repeat proteoglycans (SLRP) have emerged as a significant multi-member group, expressed in high levels in many tissues including heart and are capable of binding and regulating collagen fibril assembly and growth [28]. The four SLRP members-lumican, fibromodulin, decorin and biglycans have recently been the focus of gene targeting studies to understand their role in connective tissue biology. For example, lumican-deficient mice had collagen fibrils of increased diameters forming a disorganized matrix in the cornea and skin, with subsequent decrease in corneal clarity and increased skin laxity. Similarly, fibromodulin-deficient mice contained more immature collagens with small diameter [29]. Excluding the natriuretic factors, SLRPs constituted the most reproducible group of genes with increased expression in failing hearts. A recent microarray study by Boheler et al. [14] showed lumican as one of the only two finally passed genes overexpressed in HF after neutralizing the possible confounding effects of age, sex and data normalization procedure.

Fibronectin, a widely investigated glycoprotein in myocardial matrix, was among the consistently overexpressed genes, regardless of cause and severity of HF. In addition to the individual role of fibronectin in regulating the growth and organization of myocardial matrix, it possesses an Arginine-Glycine-Aspartate (RGD) domain that can bind to the special group of integrin molecules expressed in the heart [19]. After a broader search for the molecules sharing the RGD domain,

we noticed that more genes encoding the proteins of this family, e.g. thrombospondin-1,2,3,4, osteopontin and activin were also reported to be overexpressed in HF. Integrins are heterodimeric cell-surface adhesive receptors which bind to ECM proteins. Upon stimulation, integrins interact with signal transducing molecules like Phospho-FAK, c-src and shc [30]. Activation of these tyrosine-rich proteins leads to the phosphorylation and activation of mitogen-activated protein kinases (MAPKs) signaling cascade, finally leading to myocardial matrix growth [31]. Interestingly, MAPK was the second gene exclusively overexpressed in HF out of 10,176 clones tested by Boheler et al. The meaning for the transcriptomal increase of matrix gene expression in the context of HF can be twofold. Firstly, an increased expression of the genes involved in matrix organization can be beneficial by offering an additional structural support and resilience to failing hearts. Secondly, however, an aberrant activation of signaling cascades that are contributory to cardiac fibroblast activation and excess fibrillar collagen production can induce myocardial stiffness, lead to electromechanical uncoupling and increase myocardial ischemia, thus turning the virtuous intentions to malign consequences. Taken together, it can be concluded that, despite apparently large differences in the design of the arrays and individual experiments, the results are often complementary and converge to a common gene expression profile of failing hearts.

4. Cell-cycle regulators and protein elongation factors in failing hearts: the unusual suspects

The role of cell-cycle activation and increased mitotic activity in failing ventricles is currently a subject of interest. The active remodeling of myocardial matrix is invariably accompanied by an increased turnover of its cellular counterpart. We have observed a 10-fold increase in cyclin D1 expression in the failing hearts of Ren-2 rats compared to the littermates that had normal cardiac function [32]. A recent study by Yung et al. [18] also showed a 4.3-fold increase in cyclin G2 in failing hearts. A second interesting point raised by most of the microarray studies discussed in this review is the increased expression of protein elongation factors (eIF) in HF [6,7,10,33,34]. Increased eIF activity can lead to an increase in bulk protein synthesis, and therefore, enhance cell replication because cell division requires sufficient protein production to fulfill the metabolic requirements of the daughter cells. It was found that eIF-2 is one of the most prominently phosphorylated proteins in mammalian tissue extracts [35]. Phosphorylation of eIF-2 results in its inactivation, thus suggesting a mechanism for global protein synthesis regulation [36]. Of note, eIF-null mice develop severe cachexia and succumb to death within a month [37]. In the absence of mechanistic insight, it is probably too early to postulate that HF is a malignant form of cardiac remodeling. However, it could be argued that an unchecked cell division and protein synthesis machinery could still be the hidden suspects in failing hearts. The

therapeutic translation of this concept could lead to the testing of new cell-cycle activation inhibitors or specific inhibitors of eIF for HF therapy.

5. New publicly accessible tools to link the functionality of the genes

Since the differentially expressed genes in HF or any other disease process have either shared ontology, domain or function, there was a need for an array-clustering tool, which can map the differentially expressed genes in individual database to known biological pathways. Identifying groups of biologically related genes that are showing a large number of gene expression changes will create an informative description of the biology that is occurring in a particular dataset, making it possible to generate new hypotheses and identify those specific areas of biology that warrant more detailed investigation. One tool that assists in the identification of important biological processes is GenMAPP (Gene MicroArray Pathway Profiler) [38], a program for viewing and analyzing microarray data on microarray pathway profiles (MAPPs) representing biological pathways or any other functional grouping of genes. The second and more recently developed tool is MAPPFinder that dynamically links gene expression data to the gene ontology (GO) hierarchy [39]. MAPPFinder, therefore, generates a gene expression profile at the level of biological processes, cellular components and molecular functions, rapidly identifying those areas of biology that warrant further study. Similarly, Zeeberg et al. [40] developed GoMiner as a new tool to interpret genomic as well as proteomic data in the context of gene ontology. Using this freely downloadable tool, one can input a list of up- or downregulated genes and a list of total genes on the array, and then calculate enrichment or depletion of categories with genes that have changed expression. Other related tools like, FatiGO, OntoExpress, GoSurfer, etc. are also implemented either in web or windows version. However, the real challenge of time is to develop a strategy in which the huge database generated from large number of experiments with same hypothesis but conducted in different biological samples can be pooled and mapped. The intra- and inter-species conservation of protein sequences, similar transcriptional nature of regulatory genes and emerging knowledge about the GO should help the to develop a standard that can serve both research scientists and software developers.

6. Matricellular vs. immunological mechanisms: marker, mediator or predictor?

We have recently reported 48 genes overexpressed in failing hypertrophied hearts compared to compensated forms of cardiac hypertrophy. This study demonstrated an apparent split of the most of the overexpressed genes either as immune-related or extracellular matrix proteins. After further investi-

gation, we noticed that thrombospondin-2, a matrix-related gene was critical for the myocardial matrix integrity, since most of the null-mutant mice died of cardiac rupture after myocardial infarction or angiotensin infusion [6]. On the other hand, galectin-3 (a macrophage-derived cytokine), the gene most robustly and specifically overexpressed only in the failing hypertrophied hearts, led to cardiac fibrosis and induced HF after chronic infusion into the pericardial cavity of the healthy rats [32]. This suggests that there is a complex interplay of protective vs. causative mechanisms in failing hearts. The gene array studies performed at a particular time-point snap a mixed picture of both of these mechanisms. Therefore, it is hard to conclude, based on the expression profile data that whether a particular differentially expressed gene in question is causative for HF or not. In order to avert these difficulties, we obtained myocardial biopsies from the rat model of hypertension and followed them up to discern whether these rats will rapidly develop HF or remain compensated. Both thrombospondin-2 and galectin-3 expression was increased at an early stage of cardiac hypertrophy specifically in the group of rats that later rapidly progressed to cardiac decompensation. In contrast, several other genes including brain natriuretic peptide were not overexpressed in early stage and thus just marked the actively failing hearts. The questions addressed by these studies are unique. The existing studies have largely focused to search for the transcriptomal imprint of HF in advanced stage. However, it is important to detect the critical mediators that are involved during the transition from compensated cardiac hypertrophy to HF. An early recognition of failure-prone hearts and intervention with newer therapeutic agents might gain additional benefit over the existing treatment strategies.

Until recently, the question on what constitute the cause and epiphenomenon of HF was not answered. Several studies have now investigated the role of various myocardial matrix proteins in various animal models using gene targeting strategy [6,41]. The evidence compiles on the fact that most of the matrix-related proteins are critical for the myocardial integrity and their increased expression in HF reflects an ongoing reparative response. These studies have an advantage of leading us towards the discovery of bio-markers or predictors of HF. Despite state-of-the-art treatment, HF is still a progressive disorder characterized by high morbidity and mortality. Therefore, a new challenge of the time is to identify new causative mechanisms of HF. It is gradually becoming clear that the myocardial inflammation is a uniform accompaniment in the failing hearts. In our recent study, we found a role for early macrophage activation for the future transition to HF [32]. We have now learned more about the cytokine production, T-lymphocyte, natural killer cell and complement system activation in failing hearts. Whether monocytes are primed from the bone marrow or circulation to transmigrate to the heart and initiate HF or it is predominantly the effect of increased expression of local adhesion molecules, cytokines and leukocyte antigens thus leading to excess tissue damage and consequent scarring and loss of

myocardial function, remains to be investigated. Hopes are high for the newly discovered molecules in HF (e.g. galectin-3, gelsolin, osteopontin) that appear to activate the immune system and in the meantime, can contribute to the loss of cardiac function independent of immune activation [7,32,42].

7. Bioinformatic analysis of expressed sequence tags

With the aid of EST and microarray applications, both known and previously uncharacterized genes involved in the induction of HF can be analyzed simultaneously [43]. As an example, we chose two previously reported ESTs—ESTA, a 12.45-fold upregulated and ESTB, a 3.19-fold downregulated EST in HF [8,44]. These 2 ESTs were chosen as the index molecules for analysis based on their robust change of expression in the presence of HF. ESTA sequence was identified in the national library dbEST database with Genbank sequence ID of T60005, which corresponds to Image Clone 79419 (5') from human lung. This 329 nucleotide EST was found to be a part of Unigene cluster Hs181418 (KIAA0152 gene product), located at chromosome 12q24.31. Similarly, ESTB was annotated with the sequence ID of AA306721. This 468 nucleotide EST was found to be a part of Unigene cluster O95905 (human SGT1, hSGT1 gene product). We also confirmed the correctness of the identity of both the EST clones with their corresponding protein sequences by a direct BLASTn of the clone sequence against EMBL.

We then used the CBS signal P Prediction Server (V1.1) at the Center for Biological Sequence (<http://www.cbs.dtu.dk>) to compare the presence of signal peptide, cleavage sites, glycosylation and phosphorylation sites of these two proteins [45]. NCBI search was applied to analyze and predict the function of conserved domains. The signal peptide prediction algorithm using neural network models showed KIAA0152 as a secretory protein. Human SGT1 (hSGT1) protein, a G-protein coupled receptor-2 (GCR2) suppressor and transcriptional activator of glycolytic genes in *Saccharomyces cerevisiae*, appeared to be non-secretory, as the sequence did not harbor a signal peptide (Fig. 1A and B). KIAA0152 had six serine, three threonine and seven tyrosine phosphorylation sites whereas hSGT1 harbored 25 serine, six threonine and four tyrosine phosphorylation sites. KIAA0152 contained one glycosylation site at amino-acid position 261 and hSGT1 had two sites at position 133 and 501 respectively. NCBI conserved domain search revealed that KIAA0152 shared a homology with KOG3593, a predicted receptor-like serine/threonine kinase domain. This indicated the probable involvement of this protein in signal transduction mechanisms. Protein hSGT, on the other hand, shared a homology with KOG2406, a MADS box transcription factor domain, showing that it can potentially be active in transcription process.

It is noteworthy that a gene-by-gene approach in elucidating the genes and mechanisms involved is time-consuming and cumbersome. Using EST technology, Chien [46] have

recently generated a compendium of genes expressed in the human cardiovascular system, with the ultimate goal of assembling the intricacies of development and of disease, particularly the pathways leading to HF. Therefore, the bioinformatic analysis of both KIAA0152 and Protein hSGT should be interpreted as an example but not as a mandate for how the ESTs should be translated to full-sequence proteins and how the functionality is predicted. The second important aspect of this analysis is, notwithstanding the exciting information obtained from advanced data analysis tools, the biological experiments constitute the final and definite steps before any meaningful conclusions about the new disease mechanisms can be drawn.

8. Inter-array paradox in gene expression: biological effect or methodological heterogeneity?

There is no consensus about the optimal statistical approach for finding differences in expression among thousands of genes. Because a number of sources of systematic variation can alter expression levels especially in across-array experiments (inter-array variability), evaluation of gene differential expression can be biased. For example, the study by Tan et al. [12] compared expression profile between failing vs. functionally compensated forms of ischemic and dilated cardiomyopathy patients and found calponin as a 2.7-fold downregulated gene in HF. In contrast, the study performed by Steenman et al. [16] in comparable patient cohorts showed the opposite results. A partial explanation for this discrepancy became available when Boheler et al. [14], using general linear statistical model as a data analysis tool, attributed the altered expression of calponin to the gender-difference. Although the near-complete ascertainment of genes in the human genome should make expression-profiling studies of human hearts more powerful, identification of the sources of experimental variability, and knowledge of the relative contribution of error from each source, is critical for appropriate interpretation of the array results.

9. Chromosomal aggregation of heart failure susceptibility genes and scope for quantitative trait locus analysis

The advent of sophisticated genomic techniques for gene mapping and microarray analysis has provided opportunities to map mRNA abundance to quantitative trait loci (QTL) throughout the genome [47]. Unfortunately, simple mapping of each individual mRNA trait on the scale of a typical microarray experiment is computationally intensive, subject to high sample variance, and therefore, underpowered. However, individual susceptibility to HF could still be guided by genetic component and chromosomal mapping of the genes offer promise for understanding the molecular mechanism of the etiology and provide new therapeutic targets. As an

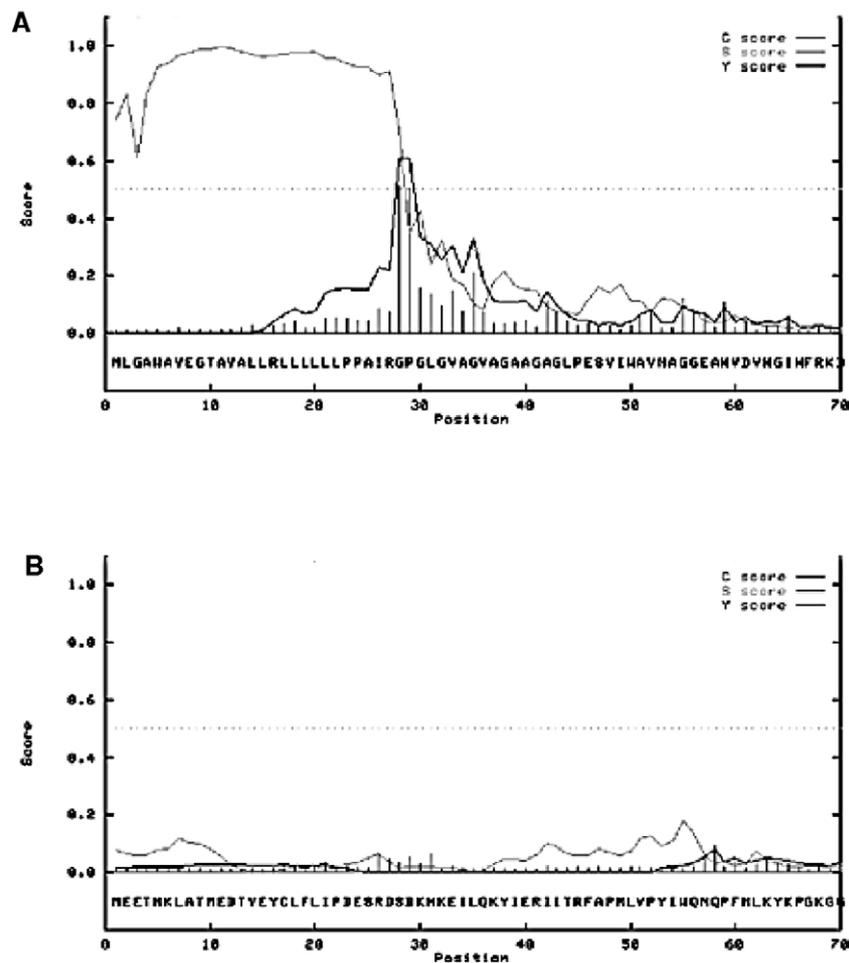


Fig. 1. Analysis of KIAA0152 and hSGT1 amino acid sequences. **A**, Signal peptide prediction plot indicates a potential leader sequence at the amino terminus of KIAA0152. Mean Signal peptide (S) score = 0.92, which represents a signal peptide consisting of the first 27 amino acids and a cleavage site between position 27 and 28. **B**, Protein hSGT1 does not harbor the leader sequence and cleavage site. C-score is the cleavage site score, which should only be significantly high at the cleavage site. Y-score is a derivative of the C-score combined with the S-score. The cleavage site is assigned from the Y-score where the slope of the S-score is steep and a significant C-score is found.

example, summarizing the chromosomal localization of the gene products that had an absolute change in expression of ≥ 1.8 -fold in multiple microarray analyses of human HF, revealed chromosome 1,2,11 and 12 as the chromosomal hotspots of HF¹⁴. A closer look on the individual genes belonging to these chromosomal loci showed several genes (e.g. natriuretic peptide precursor A and B, transforming growth factor- β , fibronectin, collagens, cathepsin C and F, phospholipase A, etc.) consistently reported to be differentially expressed in failing hearts.

10. Future directions

In the next decade, it is anticipated that the Human Genome Project and related human and animal genome sequencing projects will exert a profound influence on our understanding and management of human disease. For HF, it still remains unclear which genes are most responsible for disease onset, even though patients often report a strong familial predisposition. Expectations are high for the potential of cardiovascu-

lar genomics to lead to major advances in our understanding of normal cardiovascular functioning and pathogenesis of HF, of interactions of genes with environment, of strategies for disease prediction, and in drug development. Using sophisticated cluster analyses, novel patterns in functional responses are seen to various stimuli, and clustered responses may be useful for identifying novel pathways for design and application of therapeutics.

At first glance, the task of deciphering the results from diverse database may seem confusing. First, casting the net for genome-wide changes in gene expression will inevitably identify variations. Establishing which of this is functionally important will require better informatics and more efficient means of confirming gene function experimentally than conventional methods of genetic engineering in mice. Second, changes in mRNA need not accompany changes in the corresponding protein's abundance or its state of activation. Clearly, progress in this field will require a new level of interdependence between specialists broadly in the biological sciences and those in the population and clinical sciences. The

prospect of embarking on this voyage is at the same time daunting and exhilarating: researchers must rapidly ‘get up to speed’ on so many evolving new disciplines, but there is the opportunity to travel into uncharted territories to make discoveries at a great pace and on a large scale.

References

- [1] Yousef GM, Yacoub GM, Polymeris ME, Popalis C, Soosaipillai A, Diamandis EP. Kallikrein gene downregulation in breast cancer. *Br J Cancer* 2004;90:167–72.
- [2] Carella M, Volinia S, Gasparini P. Nanotechnologies and microchips in genetic diseases. *J Nephrol* 2003;16:597–602.
- [3] Cook SA, Rosenzweig A. DNA microarrays: implications for cardiovascular medicine. *Circ Res* 2002;91:559–64.
- [4] Napoli C, Lerman LO, Sica V, Lerman A, Tajana G, de Nigris F. Microarray analysis: a novel research tool for cardiovascular scientists and physicians. *Heart* 2003;89:597–604.
- [5] Kittleson MM, Ye SQ, Irizarry RA, Minhas KM, Edness G, Conte JV, Parmigiani GML, Chen Y, Hall JL, Garcia JG, Hare JM. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. *Circulation* 2004;110:3444–51.
- [6] Schroen B, Heymans S, Sharma U, Blankesteijn WM, Pokharel S, Cleutjens JP, Porter JG, Evelo CT, Duisters R, van Leeuwen RE, Janssen BJ, Debets JJ, Smits JF, Daemen MJ, Crijns HJ, Bronstein P, Pinto YM. Thrombospondin-2 is essential for myocardial matrix integrity: increased expression identifies failure-prone cardiac hypertrophy. *Circ Res* 2004;95:515–22.
- [7] Yang J, Moravec CS, Sussman MA, DiPaola NR, Fu D, Hawthorn L, Mitchell CA, Young JB, Francis GS, McCarthy PM, Bond M. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation* 2000;102:3046–52.
- [8] Barrans JD, Allen PD, Stamatiou D, Dzau VJ, Liew CC. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 2002;160:2035–43.
- [9] Jin H, Yang R, Awad TA, Wang F, Li W, Williams SP, Ogasawara A, Shimada B, Williams PM, de Feo G, Paoni NF. Effects of early angiotensin-converting enzyme inhibition on cardiac gene expression after acute myocardial infarction. *Circulation* 2001;103:736–42.
- [10] Barrans JD, Stamatiou D, Liew C. Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun* 2001;280:964–9.
- [11] Hwang JJ, Allen PD, Tseng GC, Lam CW, Fananapazir L, Dzau VJ, Liew CC. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* 2002;10:31–44.
- [12] Tan FL, Moravec CS, Li J, Apperson-Hansen C, McCarthy PM, Young JB, Bond M. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci USA* 2002;99:11387–92.
- [13] Blaxall BC, Tschannen-Moran BM, Milano CA, Koch WJ. Differential gene expression and genomic patient stratification following left ventricular assist device support. *J Am Coll Cardiol* 2003;41:1096–106.
- [14] Boheler KR, Volkova M, Morrell C, Garg R, Zhu Y, Margulies K, Seymour AM, Lakatta EG. Sex- and age-dependent human transcriptome variability: implications for chronic heart failure. *Proc Natl Acad Sci USA* 2003;100:2754–9.
- [15] Chen MM, Ashley EA, Deng DX, Tsalenko A, Deng A, Tabibiazar R, Ben-Dor A, Fenster B, Yang E, King JY, Fowler M, Robbins R, Johnson FL, Bruhn L, McDonagh T, Dargie H, Yakhini Z, Tsao PS, Quertermous T. Novel role for the potent endogenous inotrope apelin in human cardiac dysfunction. *Circulation* 2003;108:1432–9.
- [16] Steenman M, Chen YW, Le Cunff M, Lamirault G, Varro A, Hoffman E, Leger JJ. Transcriptomal analysis of failing and nonfailing human hearts. *Physiol Genomics* 2003;12:97–112.
- [17] Grzeskowiak R, Witt H, Drungowski M, Thermann R, Hennig S, Perrot A, Osterziel KJ, Klingbiel D, Scheid S, Spang R, Lehrach H, Ruiz P. Expression profiling of human idiopathic dilated cardiomyopathy. *Cardiovasc Res* 2003;59:400–11.
- [18] Yung CK, Halperin VL, Tomaselli GF, Winslow RL. Gene expression profiles in end-stage human idiopathic dilated cardiomyopathy: altered expression of apoptotic and cytoskeletal genes. *Genomics* 2004;83:281–97.
- [19] Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, Wang HY, Ma H, Cukerman E, Lui YQ, Gu JR, Zhang JH, Tsui SK, Wayne MM, Fung KP, Lee CY, Liew CC. A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* 1997;96:4146–203.
- [20] Tadesse MG, Ibrahim JG. A bayesian hierarchical model for the analysis of Affymetrix arrays. *Ann N Y Acad Sci* 2004;1020:41–8.
- [21] Gottardo R, Pannucci JA, Kuske CR, Bretin T. Statistical analysis of microarray data: a Bayesian approach. *Biostatistics* 2003;4:597–620.
- [22] Savoie CJ, Aburatani S, Watanabe S, Eguchi Y, Muta S, Imoto S, Miyano S, Kuhara S, Tashiro K. Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. *DNA Res* 2003;10:19–25.
- [23] Yang C, Bakshi BR, Rathman JF, Blower Jr PE. Multiscale and Bayesian approaches to data analysis in genomics high-throughput screening. *Curr Opin Drug Discov Devel* 2002;5:428–38.
- [24] Brazma A, Sarkans U, Robinson A, Vilo J, Vingron M, Hoheisel J, Fellenberg K. Microarray data representation, annotation and storage. *Adv Biochem Eng Biotechnol* 2002;77:113–39.
- [25] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
- [26] Bendall JK, Heymes C, Ratajczak P, Samuel JL. Extracellular matrix and cardiac remodeling. *Arch Mal Coeur Vaiss* 2002;95:1226–9.
- [27] Boluyt MO, O’Neill L, Meredith AL, Bing OH, Brooks WW, Conrad CH, Crow MT, Lakatta EG. Alterations in cardiac gene expression during the transition from stable hypertrophy to heart failure. Marked upregulation of genes encoding extracellular matrix components. *Circ Res* 1994;75:23–32.
- [28] Scott JE. Proteodermatan and proteokeratan sulfate (decorin, lumican/fibromodulin) proteins are horseshoe shaped. Implications for their interactions with collagen. *Biochemistry* 1996;35:8795–9.
- [29] Jepsen KJ, Wu F, Peragallo JH, Paul J, Roberts L, Ezura Y, Oldberg A, Birk DE, Chakravarti S. A syndrome of joint laxity and impaired tendon integrity in lumican- and fibromodulin-deficient mice. *J Biol Chem* 2002;277:35532–40.
- [30] Palazzo AF, Eng CH, Schlaepfer DD, Marcantonio EE, Gundersen GG. Localized stabilization of microtubules by integrin- and FAK-facilitated Rho signaling. *Science* 2004;303:836–9.
- [31] Eblen ST, Slack-Davis JK, Tarcsfalvi A, Parsons JT, Weber MJ, Catling AD. Mitogen-activated protein kinase feedback phosphorylation regulates MEK1 complex formation and activation during cellular adhesion. *Mol Cell Biol* 2004;24:2308–17.
- [32] Sharma UC, Pokharel S, van Brakel TJ, van Berlo JH, Cleutjens JP, Schroen B, Andre S, Crijns HJ, Gabius HJ, Maessen J, Pinto YM. Galectin-3 marks activated macrophages in failure-prone hypertrophied hearts and contributes to cardiac dysfunction. *Circulation* 2004;110:3121–8.
- [33] Barrans JD, Ip J, Lam CW, Hwang IL, Dzau VJ, Liew CC. Chromosomal distribution of the human cardiovascular transcriptome. *Genomics* 2003;81:519–24.

- [34] Hwang JJ, Dzau VJ, Liew CC. Genomics and the pathophysiology of heart failure. *Curr Cardiol Rep* 2001;3:198–207.
- [35] Ryazanov AG. Ca^{2+} /calmodulin-dependent phosphorylation of elongation factor 2. *FEBS Lett* 1987;214:331–4.
- [36] Ryazanov AG, Shestakova EA, Natapov PG. Phosphorylation of elongation factor 2 by EF-2 kinase affects rate of translation. *Nature* 1988;334:170–3.
- [37] Shultz LD, Sweet HO, Davisson MT, Coman DR. ‘Wasted’, a new mutant of the mouse with abnormalities characteristic to ataxia telangiectasia. *Nature* 1982;297:402–4.
- [38] Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* 2002;31:19–20.
- [39] Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol* 2003;4:R7.
- [40] Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN. Gominer: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 2003;4:R28.
- [41] Trueblood NA, Xie Z, Communal C, Sam F, Ngoy S, Liaw L, Jenkins AW, Wang J, Sawyer DB, Bing OH, Apstein CS, Colucci WS, Singh K. Exaggerated left ventricular dilation and reduced collagen deposition after myocardial infarction in mice lacking osteopontin. *Circ Res* 2001;88:1080–7.
- [42] Singh K, Sirokman G, Communal C, Robinson KG, Conrad CH, Brooks WW, Bing OH, Colucci WS. Myocardial osteopontin expression coincides with the development of heart failure. *Hypertension* 1999;33:663–70.
- [43] Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991;252:1651–6.
- [44] Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for “expressed sequence tags”. *Nat Genet* 1993;4:332–3.
- [45] Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6.
- [46] Chien KR. Stress pathways and heart failure. *Cell* 1999;98:555–8.
- [47] Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 2002;416:326–30.