Genomic Analysis of Left Ventricular Remodeling

Rizwan Sarwar, MRCP; Stuart A. Cook, PhD, MRCP

The left ventricle (LV) of the heart remodels in response to hemodynamic load through the processes of physiological or pathophysiological hypertrophy.1–3 Despite extensive study of the molecular basis of cardiac hypertrophy and LV remodeling,1,4 the mainstay of medical treatment of maladaptive remodeling remains based on therapies initially devised for treating hypertension. The limited progress in the development of specific therapies for LV remodeling has led some to refer to “the impossible task of developing a new treatment for heart failure” and is the reason why, in part, device therapies have received such recent attention.5–8 To gain a better understanding of the biology of maladaptive LV remodeling, with a view to identifying therapeutic targets, new paradigms and experimental approaches are needed, and genomics provides one such methodology.

Genomics is the study of genome-scale data sets, at the DNA or RNA level, and when combined with studies of physiological traits or disease phenotypes, genomics can be used to infer molecular insights. Unlike single-gene studies, which are the favored hypothesis-driven design,4 genomic approaches examine variation in up to gigabases of sequences to find statistical associations between transcripts and traits in a hypothesis-free system. These hypothesis-free studies have often been criticized as unfocused, but following the successes of genome-wide association studies (GWAS) and the recent achievements of integrated genetic and genomic analyses, a new era of genomic experimentation presents itself.9–14 It is notable that GWAS and integrated genomic studies in humans often identify regions associated with disease rather than specific genes and that effect sizes are small; nevertheless, these data remain highly informative.15–17 In this article, we review the genomic analysis of LV remodeling in the context of the recent past, the state of the art, and the imminent use of next-generation sequencing. A glossary of terms is provided in Table 1.

The Recent Past: Genes That Go Up or Down in the Remodeled Heart

Some 5 years after the earliest microarray-based gene expression studies of cancer,18 the first genomic analyses of cardiac remodeling and heart failure were reported.19,20 Since these initial reports, there have been many studies of gene expression in the context of concentric or eccentric remodeling and overt heart failure, and these data have been reviewed comprehensively elsewhere.21–23 The wealth of information generated from genomic analyses of LV remodeling has undoubtedly advanced our understanding of the transcriptional landscape of the LV. However, issues relating to tissue heterogeneity inherent to the heart, which are less applicable for studies of clonal diseases, may have prevented more informative insights. In addition, most studies of LV remodeling have simply reported lists of genes that go up and/or down (“elevator genomics”) and applied these data to predict prognosis, to derive new molecular insights, and for therapeutic target identification, which are disputed extensions of such studies.24–26

The questions relating to the clinical utility of genomic analysis of LV remodeling may be most relevant to the field of molecular diagnostics. Although initial studies suggested that gene expression profiles could discriminate between ischemic and nonischemic cardiomyopathy,27 a subsequent larger study has not supported this finding.28 These potential shortcomings have not diminished the publication of studies that suggest a diagnostic utility of genomic analysis of ventricular biopsies (most often of the right ventricle).29 The practicing cardiologist may point out that a coronary angiogram combined with noninvasive imaging is likely a better discriminator of disease pathogenesis and outcome.

A major driving force for genomic studies of LV remodeling is the assumed premise that transcriptional variation infers a mechanistic relationship between a change in gene expression and the remodeling process. However, transcriptional variation in response to a remodeling stimulus is time dependent and over a range of magnitudes for different transcripts.30 Furthermore, a lack of variation in a transcript does not exclude its importance for remodeling, which may be mediated at the protein level. Recent studies provide further reasons in regard to why genomic studies of dichotomized “end phenotypes” of the heart (eg, heart failure versus control) have not been overly informative.10,14,31 These new data show that conserved changes in cardiac gene expression induced by a hypertrophic stimulus are limited across genetic backgrounds and that transcript levels can be better predicted by genetic factors compared with blood pressure status (Figure 1). Furthermore, although elevated blood pressure is irrefutably associated with elevated LV mass (LVM),32 many
hemodynamic parameters correlate poorly with LVM in humans\textsuperscript{3,33,34} and the rat.\textsuperscript{35–37}

Although limited, there are compelling examples of genes that have been identified through genomic analyses of remodeling that are now targets for therapeutic intervention, and we provide 2 examples of this. The first example is that of periostin, which was shown to be markedly upregulated in the first microarray study of post–myocardial infarction (MI) remodeling.\textsuperscript{19} Subsequent studies revealed that periostin is important for post-MI remodeling and repair of the LV,\textsuperscript{38} which has led to the idea that exogenous periostin may be useful for limiting LV remodeling.\textsuperscript{39} The second example is Nix, a proapoptotic mitochondrial protein that is upregulated in genetic and pressure overload–induced models of cardiac hypertrophy in the mouse.\textsuperscript{40} Overexpression of Nix results in cardiomyopathy, whereas cardiac-specific deletion of Nix protects the heart from adverse remodeling, prevents apoptotic cell death, and preserves LV function.\textsuperscript{40,41}

### State of the Art: Combined Genetic and Correlation Analyses of Gene Expression

Rare monogenic genetic diseases that exhibit the extreme phenotype of a common disease provide invaluable insights into pathophysiological pathways and also help to direct strategies for drug discovery. The validity of this method is exemplified by the work of Brown and Goldstein,\textsuperscript{42} with the consequent development of statins. However, beyond genes for mendelian forms of hypertrophic and dilated cardiomyopathy, which currently appear to play a limited role in common forms of cardiac hypertrophy,\textsuperscript{43} our understanding of the genetic control of LV remodeling is incomplete. In addition, although GWAS have identified and replicated common variants that predispose to many common human diseases, including MI and atrial fibrillation,\textsuperscript{9} only a single GWAS of baseline LV phenotypes has been published thus far.\textsuperscript{44}

In the quest to identify new genes that regulate LV remodeling, alternative strategies are needed. One such approach is to study modifier genes in the context of a mendelian trait, which can reveal genomic regions that alter disease phenotypes in response to a single gene mutation. A good example of this is given in a recent study of genetic modifiers of hypertrophy in an extended family with the InsG791 MYBPC3 hypertrophic cardiomyopathy–causing mutation.\textsuperscript{45} In this study, 3 genomic regions regulating cardiac hypertrophy were identified, but the causative genes remain to be identified. An alternative approach, which we have applied,\textsuperscript{10,14} combines genetic mapping and gene expression for gene discovery. The success of this approach

---

**Table 1. Glossary of Terms**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele</td>
<td>One version of a pair or series of sequences at a specific genomic position</td>
</tr>
<tr>
<td>eQTL</td>
<td>Expression quantitative trait locus</td>
</tr>
<tr>
<td>Genetics</td>
<td>The study of heredity (inheritance)</td>
</tr>
<tr>
<td>Genomics</td>
<td>The study of genome level data (eg, the transcriptome or genome)</td>
</tr>
<tr>
<td>Genotype</td>
<td>Allele combination present at a specific genomic position</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of the odds</td>
</tr>
<tr>
<td>LVM</td>
<td>Left ventricular mass</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait locus; a genomic region associated with a quantitative trait or disease</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>QTT</td>
<td>Quantitative trait transcript; a method used to identify transcripts whose expression correlates with a trait</td>
</tr>
<tr>
<td>RNA-Seq</td>
<td>Sequencing of mRNA by next generation sequencing</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Intermediate trait</td>
<td>A trait (most often at the molecular level) that lies between genetic variation and a whole organ/body phenotype</td>
</tr>
</tbody>
</table>

---

**Figure 1.** Genetic regulation of cardiac gene expression is predominant over hemodynamic effects in the rat. A and B, LV gene expression profiles were generated in male rats from 3 hypertensive rat strains and from 3 normotensive control strains, and the relationships between the transcription profiles were then ascertained. A, Schematic representation of how gene expression in the 3 hypertensive strains, which exhibited varying degrees of LV hypertrophy, might be expected to be associated with LV hypertrophy. B, Unsupervised hierarchical clustering of global gene expression profiles. Gene expression profiles were predominantly determined by genetic background as opposed to blood pressure effects. Strains: Sprague-Dawley Hanover (SDH); transgenic for mouse Ren2 (TGR\textsuperscript{+/+}) on SDH background; SHR; Wistar-Kyoto (WKY), from which the SHR was derived; and Lyon (LN) hypertensive rat (LH) and Lyon low blood pressure (LL). Adapted with permission from Cerutti et al.\textsuperscript{31} Copyright © 2006, the American Physiological Society.
depends on mapping a physiological trait to the genome, determining the genetic component of gene expression, and integrating these data, which we discuss in detail.

Genetic Mapping of Phenotypes to Genomic Regions in Rodents

The essence of mapping a phenotype to the genome is to measure the phenotype in a population, correlate the phenotype with genetic markers across the genome, and identify genomic regions that are significantly associated with the phenotype.46 Broadly speaking, phenotypes/traits fall into 1 of 2 categories: qualitative or quantitative. When a specific genomic region is statistically associated with a quantitative trait, then that genomic region can be labeled as a quantitative trait locus (QTL) in which a regulatory gene(s) is most likely located (Figure 2). The strength of an association of a genetic marker with a quantitative trait identified through linkage analysis is scored by a logarithm of the odds statistic.47 In segregating populations, QTLs are genetically mapped with the use of linkage analysis compared with association testing, which is used in studies of unrelated individuals, such as in GWAS. Although the first attempts to map a QTL were made almost a century ago,48 the first QTLs were only successfully mapped in the late 1980s in maize49 and subsequently in humans.50

When linkage analysis is performed, the density of the genetic markers determines the resolution of the QTL. In simple terms, the greater the number of informative individuals studied, and the more genetic markers there are, the smaller the QTL region is, and the greater are the chances of finding the causative gene in the QTL region.51 We point out that this simplified explanation does not address the complexities of the genetic architecture underlying complex traits and that polygenic and epistatic effects within a QTL can be significant.52 Early genotyping methods, such as analyses of restriction fragment length polymorphisms53 and microsatellites,54 were labor intensive and featured low throughput. More recently, with the development of microarray technologies and the sequencing of the mouse, rat, and human genomes, commercially available genotyping platforms can now simultaneously screen hundreds of thousands of single nucleotide polymorphism markers in addition to structural variants.

Linkage analysis, although useful for studies of human families, is especially suited to studies of segregating rodent populations and was originally developed for the study of model organisms.55 In rodent studies, 2 progenitor strains that differ in a trait of interest (eg, LVM) can be bred together and the progeny genotyped and phenotyped to determine which parental genetic markers segregate with the trait. The 3 most common pedigrees used in rodent genetic studies are a backcross, an F2 intercross, and recombinant inbred (RI) strains (Figure 3).56 The backcross and F2 intercross require 2 rounds of breeding, whereas RI strains can take a decade to establish. RI strains can be maintained, in theory, perpetually and are a cumulative resource. In the rat, there are 2 large sets of RI strains, and there are 4 in the mouse (Table I in the online-only Data Supplement); these resources have generated many hundreds of QTLs. It remains the case that the genes underlying most rodent QTLs remain to be identified,
but once discovered, gene effects are often robust across species and can translate to humans. The particular utility of the rat for the study of LV phenotypes is emphasized by the 62 LVM and cardiac mass QTLs already mapped in the rat compared with only 9 in the mouse.

**Genetic Mapping, Expression Profiling, and Correlation Analysis for Disease**

**Gene Identification**

In a review of the implications of microarrays for cardiovascular medicine, we posed the following question: “To what extent is the expression profile genetically ‘hard wired’?” Around this time, this question was being addressed in elegant studies in the mouse, fly, killifish, and human lymphoblastoid cell lines. These studies consistently showed that, regardless of species or cell type, there was a significant component of genetic “hard wiring” of gene expression. After these influential studies, the idea that genetic control of gene expression might be used to help to identify the genes underlying complex traits was developed. The methodology that evolved was first termed genetical genomics and was subsequently referred to as integrative genomics and expression genetics. The approach treats gene expression as a phenotype like any other, but given the quantitative nature of the phenotype and utility of gene expression as an intermediate trait, there are significant advantages to this methodology for disease gene discovery.

To summarize the integrated genetic mapping and expression analysis approach, if a transcript’s expression level can be genetically mapped to the genome, it is termed an expression QTL (eQTL), and if this eQTL is coincident with the physical location of the eQTL gene in the genome, it is termed a cis-eQTL. As with physiological QTLs, the resolution with which an eQTL can be mapped depends on the number of informative individuals genotyped and the genetic marker density. Regions may then be further refined by haplotype analysis or single nucleotide polymorphism marker association fine mapping. To use the eQTL approach to find disease genes, it is necessary to identify cis-eQTLs that colocalize to genomic regions known to regulate disease. An eQTL that maps to regions other than its own physical location in the genome is termed a trans-acting eQTL, and these tend to occur in “hot spots.” In some instances, trans-eQTLs mapping together to a single genomic location describe a biological pathway, and this can lead to the identification of a major regulator of that pathway.

Microarray expression data can also be used for quantitative trait transcript (QTT) analysis, which correlates gene expression with quantitative traits and selects for those genes with the highest correlation as candidates for the trait. Two recent articles, published back-to-back in *Nature*, applied a QTT approach to the study of obesity. It has been suggested that QTT analysis may be used in personalized medicine; although we suggest that incorporation of genetic information would be needed to realize this aim. We point out that eQTL and QTT studies require applied statistics and bioinformatics, which are beyond the scope of this review and are discussed elsewhere.

**Examples of eQTL and QTT Studies**

The very first genome-wide eQTL study was performed in yeast, like many pioneering genomic studies, and further proof-of-principle eQTL studies followed in maize, mouse liver, and human lymphoblastoid cell lines and then in rat kidney and adipose tissue and also in mouse hematopoietic stem cells and brain. More recently, 2 studies successfully used the eQTL approach to implicated cis-acting eQTLs as candidates for childhood asthma and high-density lipoprotein cholesterol. In studies of obesity, the power of combined eQTL and QTT studies has been clearly demonstrated and revealed a network of transcripts enriched in resident macrophages in adipose tissue.

To date, there have been only a few studies of cardiac phenotypes that have taken advantage of the eQTL and QTT study design. We recently applied these methodologies to identify genes for LVM and susceptibility to heart failure.
and female mice in a region syntenic with the rat LVM QTL described in our study.\textsuperscript{10,87} It is also notable that in this mouse and also for a long-read NGS (GS 20, 454 Life Sciences). Details and full references are provided (Table III in the online-only Data Supplement).

The Imminent Future: Next Generation Sequencing of the Transcriptome

Over the last decade, oligonucleotide microarrays have provided an invaluable tool for analyzing the cardiac transcriptome and for the study of LV remodeling. However, there are significant limitations to microarray technologies that rely on hybridization of samples to predesigned sequence synthesized on a variety of support matrices.\textsuperscript{62} This methodology affords interrogation of only a fraction of a transcript, does not provide sequence information, and is semiquantitative, with limited dynamic range. In addition, only known genes and previously identified expressed sequence tags can be assessed, the number and content of which differ between platforms.

Table 2. Changing Logistics of Sequencing the Human Genome

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time taken, y</td>
<td>~13</td>
<td>~4</td>
<td>~.5</td>
</tr>
<tr>
<td>No. of scientists</td>
<td>&gt;2800</td>
<td>31</td>
<td>27</td>
</tr>
<tr>
<td>Cost of sequencing, $</td>
<td>2.7 billion</td>
<td>100 million</td>
<td>&lt;1.5 million</td>
</tr>
<tr>
<td>Coverage</td>
<td>8–10×</td>
<td>7.5×</td>
<td>7.4×</td>
</tr>
<tr>
<td>No. of institutes</td>
<td>16</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>No. of countries</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

The genomes compared in the table are those produced by the Human Genome Project and 2 other published sequenced human genomes, namely, that of Craig Venter and that of James Watson. Both the Human Genome Project and the Venter genome were sequenced with the use of the same technology, whereas the Watson genome was sequenced by the use of NGS. Reproduced from Waldman\textsuperscript{89} by permission from Macmillan Publishers Ltd. Copyright © 2008, the Nature Publishing Group.

Table 3. NGS Platforms for Transcriptome Profiling

<table>
<thead>
<tr>
<th>Company</th>
<th>454 Life Sciences</th>
<th>Dover Systems</th>
<th>Illumina</th>
<th>Applied Biosystems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current platform</td>
<td>GS FLX Titanium</td>
<td>Polonator G.007</td>
<td>Genome Analyzer II</td>
<td>SOLID 2.0</td>
</tr>
<tr>
<td>First publication of NGS transcriptomics</td>
<td>2006</td>
<td>2007</td>
<td>2008</td>
<td>2008</td>
</tr>
<tr>
<td>Read length, bp</td>
<td>Up to 500</td>
<td>~25</td>
<td>35–50</td>
<td>35–50</td>
</tr>
<tr>
<td>Data per run, Gb</td>
<td>~0.5</td>
<td>~8</td>
<td>~3</td>
<td>~6</td>
</tr>
<tr>
<td>Run time, h</td>
<td>4</td>
<td>48</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

Comparisons of the sequencing capacity and run times between NGS platforms are shown. Data per run estimates are given as approximate values and differ on a single platform depending on the application and the read length and are increasing rapidly with technological advances. Gb indicates gigabases.
Recently, next generation sequencing (NGS) technologies, which were primarily designed for DNA sequencing (Table 2) and generate up to gigabases of sequence data (Table 3), have been used for the analysis of the transcriptome (Figure 4). These platforms apply a number of sequencing technologies, including sequencing by synthesis and ligation approaches, which can be strand specific in nature. A major difference between some of the more commonly used platforms relates to the length of sequence generated (read length). This has important implications because short-read platforms (read length <50 bp) are perhaps ideally suited for quantitative digital gene expression studies (DGE; also known as Deep SAGE [DSAGE]). In contrast, longer-read platforms (read length >200 bp) may be faster, more accurate, and less bioinformatically demanding for de novo DNA sequencing and downstream genome assembly.

For transcriptome analysis, NGS affords an unbiased analysis of the transcriptional landscape and can reveal quantitative gene expression, splice variation, allele-specific gene expression, and new coding and noncoding genes and enable annotation of 3′ and 5′ untranslated regions (UTRs) of genes to within 50 bp of Rapid amplification of cDNA ends libraries (Figure 5). To date, a number of NGS studies have been performed in maize, alfalfa, yeast, and the fly, and the utility of NGS RNA-Seq for translational studies has been shown in a study of human mesothelioma. With refinements of the technology, >100× sequence coverage of more than half of expressed genes may be achieved for transcriptome studies. With the marked advantages of NGS platforms over more traditional technologies and the precipitous fall in costs associated with NGS, RNA-Seq may truly be a “revolutionary tool for transcriptomics.”

Conclusions

Numerous genomic studies of LV remodeling have been published over the last decade, and although the critic might suggest that many of these studies are observational in nature, there have been some undoubted successes. More recently, integrated eQTL and QTT approaches have revealed new genes for LV remodeling, and we anticipate further insights from this study design, GWAS, deep resequencing, and modifier gene studies over the next few years. There are some obvious “low-hanging fruit” for the next cohort of genomic studies of the heart, such as defining the cardiac transcriptome at the single-molecule resolution with the use of NGS. Although these experiments will reveal a new level of genomic information, it is important to move on from the study designs that have defined the microarray era and apply and develop cross-disciplinary approaches that combine genetics, bioinformatics, computer science, statistics, and NGS. We predict that the application of these integrated methodologies will reveal many more genes underlying cardiac QTLs and lead to the identification of critical pathways and networks for LV remodeling.

Sources of Funding

This article was supported by the Medical Research Council, UK; British Heart Foundation; and Fondation Leducq.

Disclosures

Dr Cook has a research collaboration with Applied Biosystems. Dr Sarwar reports no conflicts of interest.

References


Downloaded from http://circ.ahajournals.org/ by guest on July 18, 2017


87. Llamas B, Belanger S, Picard S, Deschepper C. Cardiac mass and cardiomyocyte size are governed by different genetic loci on either autosomes or chromosome Y in recombinant inbred mice. Physiol Genomics. 2007;31:176–182.
Genomic Analysis of Left Ventricular Remodeling
Rizwan Sarwar and Stuart A. Cook

Circulation. 2009;120:437-444
doi: 10.1161/CIRCULATIONAHA.108.797225
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circ.ahajournals.org/content/120/5/437

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2009/07/27/120.5.437.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org/subscriptions/
Supplementary Table 1 *Large sets of recombinant inbred rodent (RI) strains in rats and mice.* Details of RI strains are given, including the first publication dates. The BXD panel includes a recently generated RI intercross\(^1\). The PXO RI strains\(^2\), although not a large panel, are descended from the same progenitors as the BXH/HXB set of RI strains.\(^3\) It is predicted that the first 400 set of strains from the Collaborative Cross (CC) will be ready by 2012\(^4\), and there are plans to expand the set to include 1000 strains.\(^5\)

<table>
<thead>
<tr>
<th>Species</th>
<th>Publication</th>
<th>Progenitor strains</th>
<th>Set</th>
<th>Reciprocal</th>
<th>Number</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>1978(^1,6)</td>
<td>C57BL/6J</td>
<td>BXD</td>
<td>No</td>
<td>80</td>
<td>Jackson Lab, Bar Harbor, Maine, USA</td>
</tr>
<tr>
<td></td>
<td>1984(^7)</td>
<td>A/J, C57BL/6J</td>
<td>AXB/BXA</td>
<td>Yes</td>
<td>29</td>
<td>Jackson Lab, Bar Harbor, Maine, USA</td>
</tr>
<tr>
<td></td>
<td>1986(^8)</td>
<td>AKR, DBA/2</td>
<td>AKXD</td>
<td>No</td>
<td>24</td>
<td>Jackson Lab, Bar Harbor, Maine, USA</td>
</tr>
<tr>
<td></td>
<td>2004(^9)</td>
<td>Inbred Long Sleep (ILS) Inbred Short Sleep (ISS)</td>
<td>LXS</td>
<td>Yes</td>
<td>77</td>
<td>University of Colorado, Boulder, Colorado</td>
</tr>
<tr>
<td></td>
<td>Unpublished</td>
<td>A/J, C57BL/6J, 129S1/SvI/J, NOD/LtJ, NZO/HiLtJ, CAST/Ei, PWK/PhJ, WSB/EiJ</td>
<td>CC (in production)</td>
<td>Yes</td>
<td>~ 400</td>
<td>In production(^4,5)</td>
</tr>
<tr>
<td>Rat</td>
<td>1989(^3)</td>
<td>Brown Norway (BN.Lx/Cub) Spontaneously Hypertensive Rat (SHR/Ola)</td>
<td>BXH/HXB</td>
<td>Yes</td>
<td>32</td>
<td>Czech Academy of Sciences</td>
</tr>
<tr>
<td></td>
<td>1997(^10)</td>
<td>Fischer 344 (F) Long Evans (LE)</td>
<td>LEXF/FXLE</td>
<td>Yes</td>
<td>34</td>
<td>National Bio Resource Project for the Rat in Japan</td>
</tr>
<tr>
<td></td>
<td>2003(^2)</td>
<td>Spontaneously Hypertensive Rat (SHR.Lx) BXH2 (from BXH/HXB RI strains)</td>
<td>PXO</td>
<td>No</td>
<td>10</td>
<td>Charles University, Prague, Czech Republic</td>
</tr>
</tbody>
</table>
Supplementary Table 2 Experimental design and results of integrative genomic studies on LV remodeling. BN, Brown Norway; SHR, Spontaneously Hypertensive Rat; BP, blood pressure; LVM, left ventricular mass; HF, heart failure; SHHF, Spontaneously Hypertensive Heart Failure prone rat; SHRSP, Spontaneously Hypertensive Stroke-Prone rat; WKY, Wistar Kyoto

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BN – normal BP and LVM</td>
<td>SHR – raised BP and LVM</td>
<td>SHHF – HF and raised BP</td>
<td>SHRSP – no HF and raised BP</td>
</tr>
<tr>
<td>F2 – SHHF x SHRSP</td>
<td>F2 – SHHF x WKY</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genetic studies</th>
<th>Phenotypes</th>
<th>LVM indexed to body mass</th>
<th>Telemetric blood pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic markers</td>
<td>1,011</td>
<td>570</td>
<td></td>
</tr>
<tr>
<td>Number of rats for phenotyping</td>
<td>4-6 males/strain for LVM</td>
<td>162 rats from SHHF x SHRSP</td>
<td>189 rats from SHHF x WKY</td>
</tr>
<tr>
<td>8-12 males/strain for BP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage analysis</td>
<td>QTL on chr 17, LOD 4.0</td>
<td>QTL on chr 15, LOD 4.6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expression profiling</th>
<th>Platform</th>
<th>Affymetrix Rat 230 2.0 (31,099 transcripts)</th>
<th>Affymetrix Rat 230 2.0 (31,099 transcripts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microarrays</td>
<td>116 (4 males/RI strain)</td>
<td>177 (SHHS x SHRSP F2)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study findings</th>
<th>Summary</th>
<th>osteoglycin (Ogn) increases LVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular mechanism</td>
<td>variation in 3’UTR increases protein levels of Ogn</td>
<td>variation in promoter decreases RNA levels</td>
</tr>
<tr>
<td>soluble epoxide hydrolase (Ephx2) promotes HF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 3** *Advances in transcriptomic sequencing platforms.* This table shows further details of the data plotted in Figure 4. ‘Read length’ is defined as the length of sequence interrogated by the platform, and in the case of the Affymetrix and NimbleGen arrays, refers only to gene/EST probes on the arrays. In cases where more than one library was sequenced, the mean number of reads and sequence are per sample. SAGE, serial analysis of gene expression; Oligo, oligonucleotide; PL, photolithography; NGS, next generation sequencing.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Publication Date</th>
<th>‘Read’ length (bases)</th>
<th>‘Reads’ detected per sample</th>
<th>Total raw sequence per sample (MB)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>spotted array</td>
<td>Oct 1995¹¹</td>
<td>~ 1000</td>
<td>45</td>
<td>0.045</td>
<td>first cDNA array</td>
</tr>
<tr>
<td>SAGE</td>
<td>Oct 1995¹²</td>
<td>9</td>
<td>1,000</td>
<td>0.009</td>
<td>proof-of principle experiment</td>
</tr>
<tr>
<td>oligo array</td>
<td>Dec 1996¹³</td>
<td>25</td>
<td>16,000</td>
<td>0.40</td>
<td>Affymetrix masked PL array</td>
</tr>
<tr>
<td>SAGE</td>
<td>Jan 1997¹⁴</td>
<td>15</td>
<td>60,000</td>
<td>0.91</td>
<td>first full SAGE experiment</td>
</tr>
<tr>
<td>short read NGS</td>
<td>Jun 2000¹⁵</td>
<td>16-20</td>
<td>130,000</td>
<td>2.4</td>
<td>Lynx Therapeutics MegaClone</td>
</tr>
<tr>
<td>oligo array</td>
<td>Apr 2001¹⁶</td>
<td>60</td>
<td>24,000</td>
<td>1.4</td>
<td>Agilent ink-jet array</td>
</tr>
<tr>
<td>oligo array</td>
<td>Nov 2002¹⁷</td>
<td>24</td>
<td>65,000</td>
<td>1.6</td>
<td>NimbleGen maskless PL array</td>
</tr>
<tr>
<td>bead array</td>
<td>Nov 2004¹⁸</td>
<td>50</td>
<td>49,000</td>
<td>2.4</td>
<td>Illumina bead array</td>
</tr>
<tr>
<td>long read NGS</td>
<td>Sep 2006¹⁹</td>
<td>102</td>
<td>180,000</td>
<td>20</td>
<td>454 Life Sciences GS 20</td>
</tr>
<tr>
<td>oligo array</td>
<td>Apr 2007²⁰</td>
<td>25</td>
<td>5,500,000</td>
<td>140</td>
<td>Affymetrix exon array</td>
</tr>
<tr>
<td>short read NGS</td>
<td>Jun 2007²¹</td>
<td>14</td>
<td>2,200,000</td>
<td>31</td>
<td>Dover Systems Polonator prototype</td>
</tr>
<tr>
<td>short read NGS</td>
<td>Jun 2008²²</td>
<td>70 (2 x 35)</td>
<td>15,000,000</td>
<td>1000</td>
<td>Illumina Genome Analyzer</td>
</tr>
<tr>
<td>short read NGS</td>
<td>Jul 2008²³</td>
<td>25</td>
<td>53,000,000</td>
<td>1700</td>
<td>Applied Biosystems SOLiD</td>
</tr>
</tbody>
</table>
References


