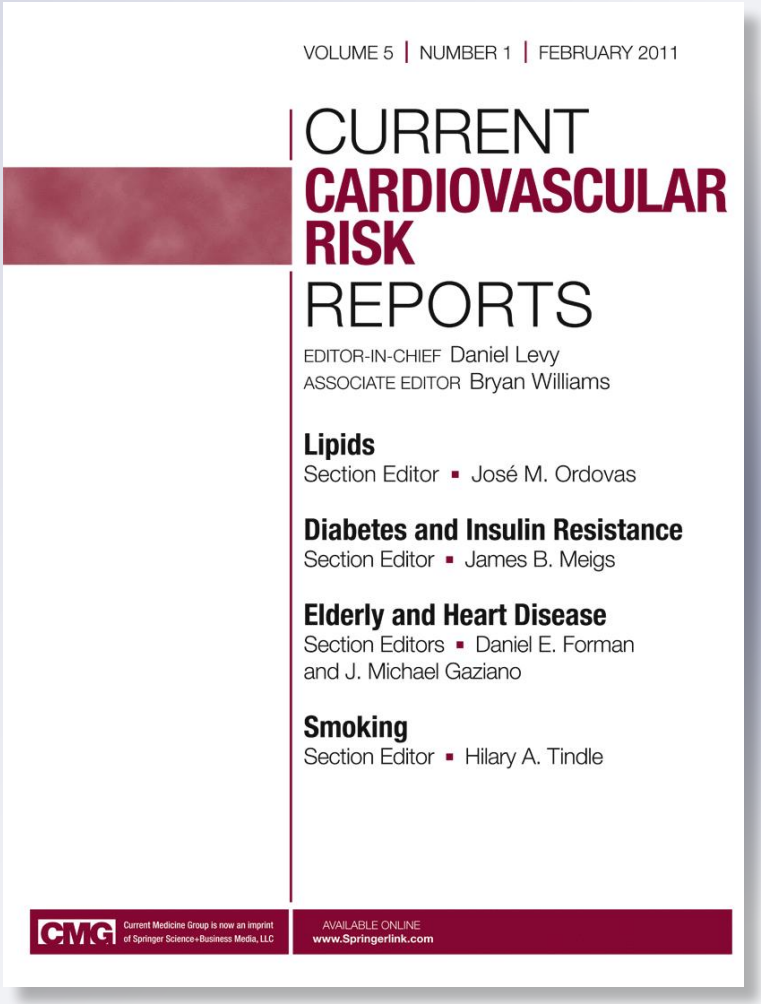
Systems Biology Approaches for Investigating the Relationship Between Lipids and Cardiovascular Disease

Current Cardiovascular Risk Reports

ISSN 1932-9520
Volume 5
Number 1

Curr Cardiovasc Risk Rep
(2010) 5:52-61
DOI 10.1007/
s12170-010-0144-2



Springer

Curr Cardiovasc Risk Rep (2011) 5:52–61
DOI 10.1007/s12170-010-0144-2

Systems Biology Approaches for Investigating the Relationship Between Lipids and Cardiovascular Disease

Gemma M. Kirwan · Diego Diez · Jesper Z. Haeggström · Susumu Goto · Craig E. Wheelock

Published online: 16 November 2010

Abstract Systems biology is an emerging field that offers promise in exploring the inter-connectivity and causality between biological pathways. This review focuses on systems biology approaches in cardiovascular disease and on the role of inflammatory lipid mediators in atherosclerosis. The basic concepts of systems biology are presented, with a focus on the integration of "omics" data from multiple technology platforms, applications of multivariate analysis, and network theory. A particular emphasis is placed on the role of multivariate statistics in analyzing data from omics platforms. An overview of selected systems biology-specific bioinformatics tools is provided, with a focus on applications that explore the role of lipids in cardiovascular systems. Systems biology offers the promise of increased insight into the biological pathways involved in cardiovascular disease and in unraveling the mechanistic relationships arising from lipid-artery interactions that lead to immune and inflammatory responses and the onset of disease.

Keywords Systems biology · Lipidomics · Cardiovascular disease · Atherosclerosis · Lipid · Eicosanoid · Inflammation · Multivariate statistics · PCA · OPLS

G. M. Kirwan · D. Diez · S. Goto · C. E. Wheelock
Bioinformatics Center, Institute for Chemical Research,
Kyoto University,
Uji, Kyoto 611-0011, Japan

J. Z. Haeggström · C. E. Wheelock (✉)
Department of Medical Biochemistry and Biophysics,
Division of Physiological Chemistry II, Karolinska Institutet,
SE-171 77, Stockholm, Sweden
e-mail: craig.wheelock@ki.se

Introduction

Systems biology is an emerging field that explores the inter-connectivity and functionality between components in a biological system in a holistic approach to understanding fundamental biological mechanisms. Organisms are complex and a biological system cannot be explained by its constituents alone, but is reliant upon system level connections to make a whole organism. Although systems-based studies are not new, increased interest is being driven by advances in molecular biology and analytical chemistry combined with concomitant increases in high-performance computing and data acquisition. Systems biology focuses on the systematic study of interactions in biological organisms, often using high-throughput "omics" technologies, such as transcriptomics, proteomics, metabolomics, and lipidomics, and aims to probe the shifts of a system not only in time, but also in distance (eg, the effects of one organ on another) [1]. A systems approach differs from a reductionist approach in that it does not focus on identifying a single gene, metabolite pathway, or mechanism, but rather examines the relationships between the components. In other words, it seeks to identify inter-connected biological networks in order to describe the shifts in equilibrium (eg, as an organism moves from a healthy to a diseased state). The large amounts of biological data often required for a systems biology study are complex and the necessary bioinformatics software and tools are under constant development. Data are often mined through a variety of analytical techniques, and the grouping and comparison of these data requires novel processing and numerical analysis methods. Once data are collated and processed, biological mechanistic model building and prediction can be performed. In this sense, the workflow of a systems biology

study combines multidisciplinary research areas beginning with clinical science, chemistry, physics, biological analysis, statistical analysis, and finally information technology and network/pathway analysis.

The holistic approach of systems biology is particularly useful for investigating complex disease such as cardiovascular disease and its sequelae [2]. Cardiovascular disease is classified as a range of afflictions concerning the heart and blood vessels, and it is multi-factorial and multi-genetic. In the United States alone, cardiovascular disease accounts for an average of one death every 37 s [3]. The most common form of cardiovascular disease is atherosclerosis, the build-up of fatty materials, fibrosis and calcification in artery walls [4]. Atherosclerosis develops from low-density lipoprotein (LDL) molecules being oxidized and deposited in the artery walls where they elicit tissue damage. This oxidative stress induces activation of pro-inflammatory gene expression, subsequent endothelial cell dysfunction, and concomitant increase in chemokine levels that promotes monocyte migration and macrophage proliferation [5]. This cascade results in a complex inflammatory and immune response, including adaptive responses, in the pathogenesis of atherosclerosis [6]. Hyperlipidemia causes a large serum enrichment of pro-inflammatory monocytes that interact with atherosclerotic lesions and propagate innate immune responses and inflammation, resulting in a mobilization of leukocytes to repair the damage. However, the leukocytes cannot effectively process the oxidized LDL and therefore grow and rupture, depositing additional LDL in the artery wall. This cycle continues, leading to inflammation of the artery wall and enlargement of endothelial cells around the artery, forming a hard sheath. Cholesterol is the major lipid species that accumulates in atherosclerotic lesions and is transported in the bloodstream by lipoproteins [7]. Cholesterol is released from lipoproteins and oxidized, leading to inflammation, which is exacerbated if insufficient high-density lipoprotein (HDL) is available. Macrophages that have accumulated oxidized LDL can form foam cells, which prompt the accumulation of more LDL. When the foam cells die, macrophage activity increases, thereby promoting the formation of fibrous capsules and increasing inflammation [8].

The uptake of large amounts of dietary cholesterol has been linked to atherosclerosis development, and recent studies show that dietary intervention can reverse carotid atherosclerosis [9]. Likewise, reduction of plasma cholesterol is the rational for medical prevention of atherosclerosis with statins, drugs that block HMG-CoA reductase and thus endogenous synthesis of cholesterol [10].The role of lipids in these processes has been extensively studied, with focus placed upon cholesterol esters and triglycerides [7]. However, there is an equally important role of inflammatory lipid mediators in the etiology of atherosclerotic plaque development [11–13]. The multi-factorial process involved in the etiology of an atherosclerotic plaque is complex and has been recalcitrant to elucidation of mechanisms via standard reductionist approaches of molecular medicine [2]. Systems biology is well-suited to explore the relationships between lipids, gene expression, immune system, and inflammatory responses involved in the etiology of cardiovascular disease. Toward this end, this article provides an overview of current methodology, multivariate statistics, network analysis tools, and software used in systems biology approaches to investigating cardiovascular disease.

Systems Biology Methodology

The workflow of a systems biology study evolves through four principle stages: sample collection, data acquisition, multivariate analysis, and network/pathway reconstruction for data integration.

Sample Collection

Although beyond the scope of this review, sample collection is obviously vital to the success of any analysis. Clinical studies require rigorous phenotyping of patients to provide clear descriptions of disease as well as establish hereditary, genetic, and environmental complications that may contribute to observed variance in data and study outcomes [14].

Data Acquisition

Systems biology approaches utilize data sets containing information and observations that can be derived from a large array of sources, including genomics, transcriptomics, proteomics, lipidomics, and metabolomics. These disparate datasets are integrated via the use of multivariate statistics and a range of informatics approaches [15].

Multivariate Analysis

Omics data are integrated using various mathematical and statistical models, which become the ‘universal language’ used to share information across datasets and analytical platforms. These models are generally built from a combination of classical univariate and multivariate statistical analysis. Traditionally, biological data sets were comprised of a few measurements from a series of samples, colloquially called “long and lean” data sets. However, modern instrumentation and computing has resulted in the acquisition of a large number of measurements (variables) from a few samples (ie, “short and fat” data sets). In order to analyze these large omics data sets, multivariate statistics

has become the tool of choice as it aims to create models that reduce multi-dimensional data into selected variables that contain the majority of the variability in the original data set, and can simultaneously highlight outliers and patterns within the data.

The first step in analyzing a large-scale omics dataset is often pre-processing of the data prior to statistical analysis. Steps that should be considered include normalization, scaling, and noise reduction. Normalization (a dynamic matrix row operation, for example dealing with different samples) and scaling (a dynamic matrix column operation, for example dealing with signals from a sample) of a data set are performed in order to ensure that the data structure is suitable for general multivariate analysis. Normalization and scaling operations serve different purposes and it is common practice to use a combination of normalization and scaling methods. Data can be normalized to total integration or unit integration (constant sum or total spectral area) or normalized to a "standard" fixed analyte (eg, lipid, protein, gene) concentration or level. Alternatively, in some cases, the concentration of a specific analyte can be determined by independent means and is subsequently used as a reference value in the analysis. To compensate for the high dynamic range in concentrations of different analytes, the normalized data can be scaled so that changes in abundant analytes do not dominate statistical models. Scaling methods include combinations of mean centering, variance scaling, Pareto scaling, and logarithmic transformation. Pareto scaling and mean centering is the most commonly used combination. A review of the various scaling methods is provided by van den Berg et al. [16], who describe how each type of data pre-treatment emphasizes different aspects of the experimental data and notes the advantages and disadvantages of each.

To test assumptions between sample populations, a hypothesis test is often employed. For example, a t-test assesses whether two groups are statistically different with respect to the variable tested. When testing significance, a risk level is used (α value, often 0.05), and the null hypothesis is rejected if the probability $P \leq \alpha$ (ie, the two populations' means are different). However, there is still a 5% probability of the observed difference being due to chance. In that case, rejecting the null produces a false positive, also called a Type I error. For large data sets, sets where thousands of hypothesis tests are performed simultaneously, this problem becomes increasingly significant. For example, if 20,000 genes are tested, then 20,000 independent t-tests are performed to assess differences between populations. Rejecting the null at $\alpha = 0.05$ produces 1000 ($20{,}000 \times 0.5 = 1000$) significant results when all the null hypotheses are true. To overcome this problem several multiple-testing correction methods have been developed, including methods to control the family-wise error rate (FWER; eg, Bonferroni method [17]) and methods to control the false discovery rate (FDR; eg, Benjamini and Hochberg [18]). FDR is a multiple-comparisons or hypothesis-testing method used to control the expected proportion of false positives (rejected null hypothesis) in a data set by setting a threshold value from the observed distributions in the data [19]. Although useful in ensuring that any reported positives are in fact true positives, there is an increased risk of simultaneously discarding true positives during the FDR correction. The issue of Type I errors inherent in omics-style datasets highlights concerns regarding the probability of a research claim being true. In reality, the probability that a research claim is true most likely depends on the power of the study, the number of other studies performed on the same research question, as well as the ratio of true to no relationships among the relationships probed in the area being investigated, as well as any inherent bias [20].

Multivariate analysis techniques can be classified into two categories—unsupervised and supervised methods. Both methods have similar goals, which center on identifying small subsets of the data that display coordinated or correlated patterns. Unsupervised methods create clusters with no pre-defined groups, and they are used for identifying patterns in the data. Supervised methods achieve similar clusters, but the user assigns which groups the variables belong to, or their dependencies, making them useful for classifying and identifying elements of the data. The most widely used, unsupervised multivariate analysis method is principal component analysis (PCA), which derives a few select principal components that describe the greatest amount of variance in the data. PCA reduces the dimensionality of data while at the same time accounting for as much of the variation in the original data as possible. PCA transforms data to a new set of coordinates or variables, the principal components (PCs) that are linear combinations of the original variables. This is done by rotating and transforming the original variables' axes so that the new axes lie along the direction of maximum variance of the data. Interpreting and understanding of the data can be gained by examining the observations in the new factor space. PCs are perpendicular (uncorrelated) and are composed of "scores" and "loadings". Scores contain information about the observations (eg, clinical subjects), and loadings contain information about the variables (eg, lipids). PCA can be used to examine underlying features of a data set, including interactions and variable groupings.

Supervised multivariate analysis methods such as partial least-squares and/or projection to latent structures (PLS) are common, and the user defines the variables that belong to the control and response data sets prior to analysis. PLS is a technique combining features of PCA and regression. It aims to construct predictive models when a large number of

samples, observations, or variables have been recorded and are highly correlated. In PLS, observations are stored in one matrix and corresponding predictors stored in a separate matrix. PLS is a regression model that can be used to predict classifications from observations, enabling simultaneous multiple regression between the variables in the two matrices. The term "partial least-squares" describes the computation of the optimal least-squares fit to part of a covariance or correlation matrix between external variables and the dependent measures made from observation variables. PLS measures the covariation between two or more blocks of variables and creates a new set, optimized for maximum covariance (not correlation) using the fewest dimensions; PLS decomposes both predictor and observation matrices as a product of a common set of orthogonal factors and a set of specific loadings. The predictor and observation scores are chosen so that the relationship between successive pairs of scores (similar to PCA scores) is as strong as possible. PLS discriminate analysis (PLS-DA) is similar to PLS except the response variables are replaced with dummy variables that define the categories for predictions. In this manner responses are defined only by user-defined or desired classifications. Structured noise (orthogonal variation) in data causes problems for projection based methods such as PLS.

Orthogonal PLS (OPLS) is an extension of PLS, featuring an integrated orthogonal signal correction (OSC) filter [21]. Orthogonal signal correction is a pre-processing method used to filter and remove uncorrelated (orthogonal) variation within a data set X that is not correlated with Y. This orthogonal component in OPLS removes variance that is not correlated to the user matrix and represents it as variance in the X (observations) matrix. OPLS uses the information in the user (predictor) matrix to decompose the observational matrix into correlated orthogonal and residual structures of information, respectively. OPLS can find predictive components that simultaneously maximize the covariance and correlation between X and Y or observations and predictors. OPLS-DA, similar to PLS-DA, is used to find the relationship between descriptors and class identities described by a dummy matrix containing class information.

A common problem in multivariate analysis is choosing the number of components (eg, PCs) nominally required to describe the original data. Choosing too few components will result in information loss and choosing too many will include noise and/or result in overfitting. The simplest method for choosing the number of components in a multivariate analysis is still visually. For example, in PCA, a Scree Plot can be examined that plots the eigenvalues graphically with respect to the number of PCs. The optimum number of PCs is the point at which the smooth decrease of eigenvalues appears to level off to the right of the plot. A common, more robust, method to estimate the number of components required is internal cross validation. Internal cross validation is performed by removing random elements of the data matrix, then creating a model out of the remaining data using the multivariate algorithm of choice, and predicting the removed data from the new model. This is done a set number of times and the predicted data sets are then combined and compared to the original data and the sum of squared errors calculated for the whole data set (This is the predicted sum of squares [PRESS] method). The Q^2 value is then calculated from 1—PRESS / SSX where SSX is the initial sum of least squares value for the data. The maximum value for Q^2 is 1.0, which means that all variance is represented in the predictions. Good predictions have low PRESS and high Q^2 values [22].

The application of integrating these different multivariate approaches was demonstrated in a recent study that employed OPLS and PLS-DA to examine the plasma metabolome of atherosclerosis patients [23•]. Gas chromatography-mass spectrometry (GC-MS) and ^{1}H nuclear magnetic resonance (NMR) metabolomics data for patients and matched controls were initially clustered using PLS-DA and OPLS, which gave a strong OPLS model (R^2Y and Q^2 values of 1.000 and 0.435, respectively) using disease as the Y variable. The results of the loadings plots were then used to explore correlations between the metabolites identified to be responsible for separation of the two groups. This approach identified metabolites involved in insulin resistance as being relevant to the generated model, which was then visualized in a network fashion using the OPLS values. Accordingly, multivariate analysis is a vital tool in a systems biology workflow [24], but interpretation requires skill. In particular, as the data dimensionality is reduced in the analysis, it can be challenging to interpret which variables are significant in a large data set.

Network and Pathway Analysis for Data Integration

A network is defined as a series of points (nodes) interconnected by communication paths. Networks formalize the interactions between different components in a system, using integrated number theory and information technology theories [25]. Networks can be compartmentalized into nodes, which are elements of the system, and edges, which are the relationships that exist between nodes. In a metabolic network, nodes can be enzymes and compounds and edges chemical reactions and interactions, and in a protein-protein interaction network, two nodes connected by an edge represent protein interaction. The network analysis of components allows for a global overview of the system under study, achieved through statistical and

numerical analysis. The importance of individual components is not lost during the network analysis of a system as a whole. Mathematical models created through network analysis can predict the behavior of the network when components are altered or removed. An example is a network where genes form nodes—if a mutation knocks out a gene expression, the stability of the network and implications can be modeled.

Networks can be constructed from raw data derived from a variety of sources (transcriptomics, proteomics, metabolomics, lipidomics, etc), which are in turn comprised of information from a wide array of instrumental and analytical sources. The construction of the network is a critical step that is performed through the utilization of different mathematical techniques, including but not limited to, Boolean and Bayesian networks. Boolean networks use a logic system with only two variable states, 0 and 1, or "on" and "off." Originally designed for analyzing genetic regulatory systems, Boolean networks are simple and effective. The expression of each gene in the network is functionally related to the expression states of other genes, and a change in one variable (ie, a gene expression changing from 0 to 1) can be examined for the widespread changes it may cause in other variable states throughout the network. Bayesian networks are composed of defined mathematical probabilities and conditional dependencies, and are known as causal probabilistic networks. A classic example is a Bayesian network of symptoms and diseases; where, given symptoms, the network could predict the probability of various diseases [2, 26].

Models are often applied to a single network (eg, the co-transcription network), and although relationships can be modeled within a specific network, it is important to remember that an organism is not an isolated system and that any network will inter-connect with the complete system. It is therefore necessary to integrate cross-discipline networks and studies to understand systems as a whole, and work on network integration is still progressing [27•].

Software capable of analyzing omics data from various platforms and performing multivariate and numerical analysis is undergoing continual development. Some of the more common packages are listed in Table 1. There are also a number of lipid-specific databases and software platforms (Table 2), notably LIPID MAPS and LipidBank. The LIPID MAPS consortium has provided a comprehensive lipid classification, nomenclature, and structural representation system suitable for bioinformatics databasing required to analyze the numerous molecular species of lipids. The LIPID MAPS gateway includes a number of online tools for drawing lipid structures, structure prediction from mass spectral data, and a proteome database containing lipid-specific protein sequences as well as method and analytical notes. LipidBank contains information on a large variety of identified lipid structures (over 7,000 molecules classified into 26 groups), including fatty acids, steroids, glycerolipids, and sphingolipids. Molecular structures, IUPAC and common names, IR, UV, mass spectrometry, and NMR reference spectra are all available as well as links to the literature sources that identified and characterized the lipid. Taken together, these two resources provide a number of important first steps toward a comprehensive resource for the analysis of lipids, but there is still a distinct need for lipid-specific pathway reconstruction tools.

A useful resource is the Kyoto Encyclopedia of Genes and Genomes (KEGG), which is an Internet-based resource linking a series of biological databases containing information about genes, proteins, chemical building blocks, and maps of molecular interactions and biological relationships. KEGG is able to visualize nodes and edges into pathway maps for interpretation (KEGG PATHWAY). The KEGG PATHWAY suite includes >1,300 individual lipid species and their relevant information for pathway analysis. KEGG has recently developed the KegArray tool, which is a Java (Oracle, Redwood Shores, CA) application that provides an environment for analyzing transcriptomics, proteomics and/or metabolomics datasets. An application of the KegArray tool for the analysis of cardiovascular disease is shown in Fig. 1. The integration of omics data sets into KEGG databases is still problematic because there is no standard data-exchange format used in omics technology platforms. To overcome these issues, data repositories can be annotated to provide improved context [28]. Annotated pathway information can then be used to assist computer programs with managing semantics and inquiries when developing and analyzing networks.

Systems Biology Approaches in Cardiovascular Disease

The application of systems biology to investigating cardiovascular disease is still in its infancy; however, a large amount of literature is emerging. One of the earliest metabonomics studies was conducted by Brindle et al. [29], who analyzed human serum ^{1}H NMR spectra using PLS-DA to diagnose the presence of cardiovascular disease. Their models predicted the presence of coronary disease with 92% accuracy (based on a 99% CI). The most influential factors attributed to separation of diseased and healthy patients was the presence of low-density lipoprotein (LDL) and very low-density lipoprotein (VLDL) [29]. Kirschenlohr et al. [30] re-examined the heart disease predictor ability of ^{1}H NMR plasma spectra and found that only 80.3% of patients not receiving statin therapy and only 61.3% of patients treated with statins were identified. The differences in study findings were attributed to variation in

Table 1 Selected systems biology and network analysis software

Software	Description
KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.genome.jp/)	An Internet-based resource that contains a series of databases of biological systems, consisting of KEGG GENES, KEGG LIGAND, KEGG PATHWAY, and KEGG BRITE
PathViso (http://www.pathvisio.org)	A tool for displaying and editing biological pathways
WikiPathways (http://www.wikipathways.org)	An open, public platform comprised of graphical pathway editing tools and integrated databases
pSTIING (Cladist) (http://pstiing.licr.org/software/)	An Internet-based application containing metabolic pathways, protein molecule interactions, and transcriptional association.
MetaCore (http://www.genego.com/metacore.php)	A commercial software application that is designed for functional analysis of experimental data
Cytoscape (http://www.cytoscape.org)	An open source platform for visualizing and analyzing different types of networks, including biological and social networks
VANTED (http://vanted.ipk-gatersleben.de/)	A multiplatform tool for the manipulation of pathway visualization
Pathway-Express (http://vortex.cs.wayne.edu/projects.htm#Pathway-Express)	An automated profiling program for microarray data
R spider (http://mips.helmholtz-muenchen.de/proj/rspider)	An Internet-based tool for the analysis of a gene list using Reactome and KEGG databases
Bioconductor (http://www.bioconductor.org/)	An open source project for the analysis of microarray, sequence data, and high throughput assays
igraph (http://igraph.sourceforge.net/index.html)	A free software package that implements algorithms for network analysis methods, such as community structure search. It supports multiple file formats and programming languages.
Vizent Digital (http://www.vizentdigital.com/en)	An information technology solutions company focusing on pharmaceutical custom databases and software implementation

drug regimens (the use of statins), which significantly affect LDL levels and became discriminating factors in Brindle et al.'s PLS-DA analysis. Both Kirschenlohr et al. [30] and Brindel et al. [29] highlight the potential use of blood plasma–based ^{1}H NMR metabonomics and demonstrate the necessity of examining individual lipid species in cardiovascular disease diagnosis.

Some of the first studies integrating multiple omics platforms in cardiovascular research were conducted by groups at Beyond Genomics [31–33], who examined changes in the apolipoprotein E3-Leiden (apoE) transgenic mouse model. Novel data processing and multivariate analysis tool were used to compare protein and metabolic data, capturing snapshots of metabolic and protein states, confirming the function of the apoE3 gene as a lipoprotein mediator. In a separate study, de Roos et al. [34] also used the apoE knockout mouse to investigate the link between conjugated linoleic acids and atherosclerosis mechanisms. The results drawn from physiological and protein analysis, coupled with correlation analysis, suggested that consump-

Table 2 Lipid-specific bioinformatics resources

Software	Description
LIPID MAPS (http://www.lipidmaps.org)	A relational database encompassing structure and annotations of biologically relevant lipids
Lipidbank (http://lipidbank.jp/)	Official database of the Japanese Conference on the Biochemistry of Lipids
MassBank (http://www.massbank.jp)	High-resolution mass spectral database
Cyberlipid (http://www.cyberlipid.org/)	Classification scheme and extensive tools
The Lipid Library (http://www.lipidlibrary.co.uk)	Extensive, highly referenced information
Lipid Navigator (http://lipidsearch.jp/)	Identification tool for phospholipid batch processing
LIPIDAT (http://www.lipidat.tcd.ie/)	Lipid thermodynamic and phase transition information
International Union of Pure and Applied Chemistry (IUPAC) (http://www.chem.qmul.ac.uk/iupac/lipid/)	IUPAC lipid nomenclature

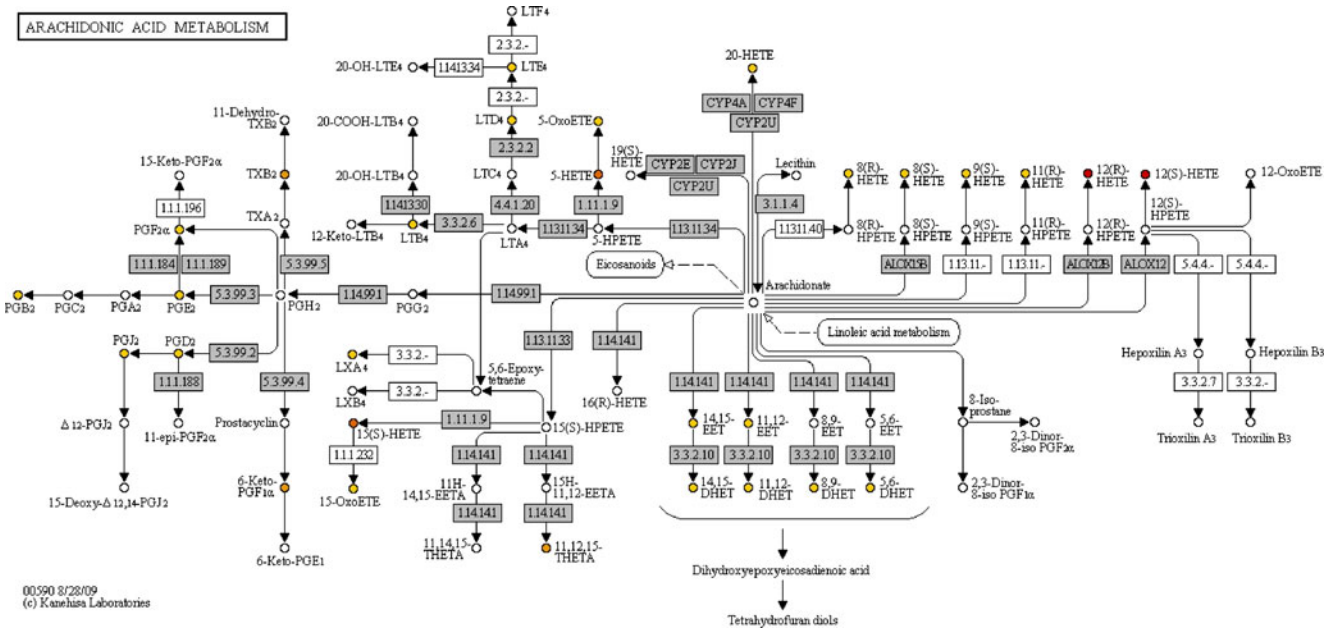


Fig. 1 KegArray-based analysis of inflammatory lipid mediators (oxylipins) from atherosclerotic plaques from the carotid artery. The resulting highest abundance KEGG metabolic pathway was "arachidonic acid metabolism" (hsa00590) and is displayed with the 27 mapped oxylipins (eicosanoids). Chirality was not determined in this study, and accordingly both isomers of 8-(*R*/*S*)-HETE and 12-(*R*/*S*)-HETE were mapped to KEGG giving 29 compounds; however the relative levels of the individual enantiomers are unknown. Data are presented as percent composition of all oxylipins quantified in the plaque (n=61) with the colors red, orange, and yellow representing> 0.8%, 0.3% to 0.4%, and<0.3%, respectively. (*Data from* Gertow et al. [50])

tion of *cis*-9- and *trans*-11-CLA increased expression of anti-inflammatory HSP 70, which could protect against atherosclerosis. Cheng et al. [35] utilized the apoE mouse in a comparative study against an LDL receptor–null mouse and found that a high-fat diet showed choline metabolism perturbations in the null mouse similar to those observed in the apoE mouse. Consistent with previous studies, both mice developed hypercholesterolemia and atherosclerosis when fed a high-fat and high-cholesterol diet.

Pietiläinen et al. [36] used a systems biology approach to determine the effects of obesity on lipid metabolism. Obesity caused metabolic alterations that promoted inflammation, atherosclerosis, and insulin resistance. One of the earlier system biology–like studies on coronary artery segments was performed by King et al. [37], who found that inflammation is more prominent in diabetic coronary disease. Microarray gene expression profiles were combined with the published literature using the Cytoscape software package to construct a network, providing insight into links between severe coronary artery disease and inflammatory and immune responses in diabetes patients. Lipopolysaccharides (LPS) are a known inducer of sepsis, which is characterized by inflammation. Tseng et al. [38] attempted to develop diagnostic markers for inflammation by examining molecular dynamics of LPS stimulation on rat endothelial cells. Bulk gene search system for Java (BGSSJ) and KEGG software was used to map differently expressed genes. Results showed that inflammatory pathways were significantly affected by LPS stimulation via NF-κB–associated responses in endothelial cells.

Statins have demonstrated efficiency in lowering cholesterol levels; however, there is concern over statin-induced myopathy after aggressive treatment. Laaksonen et al. [39] examined cellular mechanisms leading to myopathy using a systems biology approach. Results showed that the arachidonate 5-lipoxygenase activating protein gene (ALOX5AP) could act as a biomarker for statin induced myopathy [39]. The ALOX5AP gene has also been shown to predispose humans to atherosclerosis [11], with the 5-LO (5-lipoxygenase) pathway indicated to have a role in inflammatory responses during atherosclerosis [13].

Skogsberg et al. [40••] used transcriptional profiling of atherosclerosis-prone mice with human-like hypercholesterolemia to reverse engineer a network of cholesterol-response atherosclerosis target genes. Kleemann et al. [41] studied the effect of dietary intake of cholesterol upon atherosclerosis. They demonstrated that the liver is capable of removing up to approximately 0.5% w/w dietary cholesterol before the expression of hepatic pro-inflammatory and pro-atherosclerosis genes is evoked. Network analysis showed that inflammatory pathways and lipid metabolism are closely linked via specific transcriptional regulators [41]. The data from the Kleemann et al.

[41] study were further analyzed using the KegArray software [42], with distinct diet-based shifts observed in multiple KEGG pathways, especially "biosynthesis of unsaturated fatty acids" and "sphingolipids." Data suggested that even a low dose of cholesterol significantly affects the biochemical pathways responsible for lipid handling.

Van Erk et al. [43••] recently investigated the link between chronic low-level inflammation in obese subjects and health complications, including cardiovascular disease and diabetes. Over 300 plasma metabolites, including lipids, oxylipins, and fatty acids, were examined and data subjected to multivariate and correlation analysis. Novel markers of inflammatory modulation were speculated to be peripheral blood mononuclear cell (PBMC) expression of annexin A1 and caspase 8, and the arachidonic acid metabolite 5,6-DHET [43••]. Inouye et al. [44••] conducted a large-scale investigation into mechanisms of atherosclerosis using blood extracts from 500 unrelated individuals. Correlations between gene expressions and lipid concentrations demonstrated new inflammatory response pathways involving basophil granulocytes and mast cells. Network analysis determined that a previously unknown gene module, lipid leukocyte (LL), is partially reactive to blood lipid levels, giving new insight into inflammatory response mechanisms.

Lipid Mediators in Cardiovascular Disease

Changes in the homeostatic control of lipid metabolism have been acknowledged to play important roles in the pathogenesis of common diseases including cancer, insulin-resistant diabetes, and atherosclerosis. The role of lipids in atherosclerosis development is well understood, and elevated cholesterol serum levels correlate with increased atherosclerotic disease [7], but many different lipid types have been studied and linked to atherosclerosis. There are a number of mediators of the inflammatory component of cardiovascular disease, but the eicosanoids have been studied in detail for their contribution to the onset of disease [11, 12, 45, 46]. Eicosanoids are oxygenated metabolites of arachidonic acid that are formed via oxidization by three primary enzymatic pathways to a range of inflammatory lipid mediators: cyclooxygenase (COX), lipoxygenase (LOX), and cytochrome P450 (CYP450) [47, 48•]. The eicosanoids have a range of diverse biological activities in multiple diseases, with distinct roles in cardiovascular disease. For example, it has been shown that atherosclerosis correlated with high levels of prostanoids and the enzymes responsible for their formation, COX-1 and COX-2, suggesting that COX inhibitors may modulate atherosclerosis [49]. In addition, the expression of 5-LOX and leukotriene A4 hydrolase in human atherosclerotic lesions was shown to correlate with symptoms of plaque instability [11]. Accordingly, there are data indicating a role for eicosanoids in the etiology and pathology of cardiovascular disease. However, the pathways involved and their potential systemic effects are convoluted, making this area an interesting target for systems-based studies. The applications of systems biology to investigating the contribution of eicosanoids and other lipid mediators in cardiovascular disease is still in its infancy, but early work appears promising [43••]. We recently examined the levels of eicosanoids and other oxylipins in human carotid atherosclerotic lesions [50], and subsequent data were mapped to the KEGG PATHWAY database using the KegArray tool. Over 60 individual lipid mediators were quantified in the plaques from multiple polyunsaturated fatty acid (PUFA) pathways. In total, the linoleate-derived species represented >99% of the included oxylipins. These results showed that multiple oxylipin-specific metabolic pathways are present in carotid lesions. However, although only representing <1% of the total oxylipins in the plaques, arachidonic acid–derived oxylipins (eicosanoids) are well known for their role in cardiovascular disease. Accordingly, the 27 distinct arachidonic acid metabolites identified were mapped to the KEGG PATHWAY (Fig. 1). These data show that there is distinct 15-LOX and 5-LOX activity in the plaque, as demonstrated by their enzymatic products (15-HETE and 5-HETE, respectively). The utility of the KegArray tool is demonstrated by its ability to quickly provide a graphical display of many of the quantified lipid mediators; however, a major limitation is that a significant number of these compounds are not yet present in KEGG, demonstrating one of the challenges in lipid-mediator pathway mapping. These data clearly demonstrate that cells within plaques are capable of biosynthesizing a number of key inflammatory lipid mediators, and further efforts should be made to examine the production and role of these compounds in the pathophysiology of atherosclerosis.

Conclusions

Systems biology analyses involve the integration of large-scale biological datasets derived from multiple analytical methods. Datasets from different omics platforms (eg, transcriptomics, proteomics, metabolomics, and lipidomics) differ in their layout, structure, and informational storage parameters. Accordingly, the combination and integration of these datasets present one of the main challenges in systems biology and will require modern computing solutions and large teams of individuals each with disparate skills. There is a particular need for expertise in the

development of multivariate statistical approaches to deal with the unique data structures of omics-based datasets. The traditional paradigm of univariate-based statistical analysis is not appropriate, yet there is a potential for the loss of data when standard FDR methods are applied. Accordingly, integrative approaches designed for analyzing entire datasets such as OPLS-DA will become increasingly useful in systems biology studies. Systems biology offers the potential of new insights into how organisms function. For example, in the study of cardiovascular diseases, systems biology approaches are beginning to piece together and explain the complex interactions that occur between environmental stimuli (eg, diet, pollutants, lifestyle), gene expression, and lipid metabolism as well as inflammatory and immune responses on overall human phenotype. Lipid metabolism is a known factor in atherosclerosis development. Originally, atherosclerosis was believed to be caused by LDL accumulation on the artery wall, although now complex inflammation responses and lipid metabolism mediators are acknowledged as having a vital role in disease onset. The next logical step in investigations is to apply integrative systems-based approaches to these larger datasets, with the inclusion of qualitative patient meta-data in combination with quantitative variables in order to build a model of disease. The final goal is to provide an understanding of the complex interplay between the systems that constitute an organism.

Acknowledgments This research was supported by VINNOVA and JSPS under the Sweden-Japan Research Cooperative Program, the Center for Allergy Research, the Åke Wibergs Stiftelse, the Jeanssons Stiftelse, and the Swedish Research Council. GMK was supported by a JSPS Postdoctoral Fellowship, and CEW was supported by a research fellowship from the Center for Allergy Research.

Disclosure Gemma M. Kirwan reports no potential conflict of interest relevant to this article. Diego Diez reports no potential conflict of interest relevant to this article. Jesper Z. Haeggström reports no potential conflict of interest relevant to this article. Susumu Goto reports no potential conflict of interest relevant to this article. Craig E. Wheelock reports no potential conflict of interest relevant to this article.

References

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. van der Greef J, Martin S, Juhasz P et al.: The art and practice of systems biology in medicine: mapping patterns of relationships. J Proteome Res 2007, 6: 1540-59.
2. Lusis AJ, Weiss JN: Cardiovascular Networks: Systems-Based Approaches to Cardiovascular Disease. Circulation 2010, 121: 157-170.
3. Rosamond W, Flegal K, Furie K et al.: Heart Disease and Stroke Statistics 2008 Update: A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2008, 117: e25-e146.
4. Libby P: Inflammation in atherosclerosis. Nature 2002, 420: 868-874.
5. Lusis A, Attie A, Reue K: Metabolic syndrome: from epidemiology to systems biology. Nature Reviews Genetics 2008, 9: 819-830.
6. Hansson G, Libby P: The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 2006, 6: 508-519.
7. Maxfield F, Tabas I: Role of cholesterol and lipid organization in disease. Nature 2005, 438: 612-621.
8. Li A, Glass C: The macrophage foam cell as a target for therapeutic intervention. Nat Med. 2002, 8: 1235-1242.
9. Shai I, Spence J, Schwarzfuchs D et al.: Dietary intervention to reverse carotid atherosclerosis. Circulation 2010, 121: 1200-1208.
10. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 1994, 344: 1383-9.
11. Qiu H, Gabrielsen A, Agardh H et al.: Expression of 5-lipoxygenase and leukotriene A4 hydrolase in human atherosclerotic lesions correlates with symptoms of plaque instability Proc. Natl. Acad. Sci. U. S. A. 2006, 103: 8161-8166.
12. Praticò D, Dogné J: Vascular biology of eicosanoids and atherogenesis. Expert Rev Cardiovasc Ther. 2009, 7: 1079-1089.
13. Poeckel D, Funk C: The 5-lipoxygenase/leukotriene pathway in preclinical models of cardiovascular disease. Cardiovasc Res. 2010, 86: 243-253.
14. Chorro F, Such-Belenguer L, López-Merino V: Animal models of cardiovascular disease. Rev Esp Cardiol. 2009, 62: 69-84.
15. Joyce AR, Palsson BO: The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol 2006, 7: 198-210.
16. van den Berg R, Hoefsloot H, Westerhuis J et al.: Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 2006, 7: 142.
17. Bonferroni C, *Teoria statistica delle classi e calcolo delle probabilit*. 1936, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. p. 3-62.
18. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 1995, 57: 289-300.
19. Higdon R, van Belle G, Kolker E: A note on the false discovery rate and inconsistent comparisons between experiments. Bioinformatics 2008, 24: 1225-1228.
20. Ioannidis J: Why most published research findings are false. PLoS Medicine 2005, 2: e124.
21. Trygg J, Wold S: Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics 2002, 16: 119-128.
22. Stacklies W, Redestig H, Scholz M et al.: pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 2007, 23: 1164-1167.
23. • Teul J, Rupérez F, Garcia A et al.: Improving metabolite knowledge in stable atherosclerosis patients by association and correlation of GC-MS and 1H NMR fingerprints. J Proteome Res. 2009, 8: 5580-5589. *The authors demonstrate good use of multivariate analysis to explore correlations between plasma metabolites in healthy subjects and patients with carotid atherosclerosis.*
24. Madsen R, Lundstedt T, Trygg J: Chemometrics in metabolomics–a review in human disease diagnosis. Anal Chim Acta 2010, 659: 23-33.
25. Barabasi AL, Oltvai ZN: Network biology: understanding the cell's functional organization. Nat Rev Genet 2004, 5: 101-13.
26. Schadt EE, Friend SH, Shaywitz DA: A network view of disease and compound screening. Nat Rev Drug Discov. 2009, 8: 286-295.

27. • Diez D, Wheelock A, Goto S et al.: The use of network analyses for elucidating mechanisms in cardiovascular disease. Mol Biosyst. 2010, 6: 289-304. *This reference provides a solid overview of systems biology and network analysis (in the context of cardiovascular disease), including an example network constructed using microarray data from carotid endarterectomies from the Karolinska University Hospital.*
28. Kelder T, Conklin B, Evelo C, Pico A: Finding the right questions: exploratory pathway analysis to enhance biological discovery in large datasets. PLoS One 2010, 8: e10000472.
29. Brindle J, Antti H, Holmes E et al.: Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med. 2002, 8: 1439-1444.
30. Kirschenlohr H, Griffin J, Clarke S et al.: Proton NMR analysis of plasma is a weak predictor of coronary artery disease. Nat Med. 2006, 12: 705-710.
31. Orešič M, Clish C, Davidov E et al.: Phenotype characterisation using integrated gene transcript, protein and metabolite profiling. Appl Bioinformatics 2004, 3: 205-217.
32. Davidov E, Clish C, Oresic M et al.: Methods for the differential integrative omic analysis of plasma from a transgenic disease animal model. OMICS 2004, 8: 267-288.
33. Clish C, Davidov E, Oresic M et al.: Integrative biological analysis of the APOE*3-leiden transgenic mouse. OMICS 2004, 8: 3-13.
34. de Roos B, Rucklidge G, Reid M et al.: Divergent mechanisms of cis9, trans11- and trans10, cis12-conjugated linoleic acid affecting insulin resistance and inflammation in apolipoprotein E knockout mice: a proteomics approach. FASEB J 2005, 19: 1746-1748.
35. Cheng K, Benson G, Grimsditch D et al.: A metabolomic study of the LDL receptor null mouse fed a high-fat diet reveals profound perturbations in choline metabolism that are shared with ApoE null mice. Physiol Genomics 2010, 41: 224-231.
36. Pietiläinen K, Sysi-Aho M, Rissanen A et al.: Acquired Obesity Is Associated with Changes in the Serum Lipidomic Profile Independent of Genetic Effects—A Monozygotic Twin Study. PLoS One 2007, 2: e218.
37. King JY FR, Tabibiazar R, Spin JM, Chen MM, Kuchinsky A, Vailaya A, Kincaid R, Tsalenko A, Deng DX, Connolly A, Zhang P, Yang E, Watt C, Yakhini Z, Ben-Dor A, Adler A, Bruhn L, Tsao P, Quertermous T, Ashley EA.: Pathway analysis of coronary atherosclerosis. Physiol Genomics 2005, 23: 103-118.
38. Tseng H, Juan H, Huang H et al.: Lipopolysaccharide-stimulated responses in rat aortic endothelial cells by a systems biology approach. Proteomics 2006, 6: 5915-5928.
39. Laaksonen R, Katajamaa M, Päivä H et al.: A Systems Biology Strategy Reveals Biological Pathways and Plasma Biomarker Candidates for Potentially Toxic Statin-Induced Changes in Muscle. PLoS One 2006, 1: e97.
40. •• Skogsberg J, Lundström J, Kovacs A et al.: Transcriptional Profiling Uncovers a Network of Cholesterol-Responsive Atherosclerosis Target Genes. PLoS Genetics 2008, 4: e1000036. *The authors examined the progression of atherosclerosis lesions in mice and discovered a regulatory gene network centered around poliovirus receptor-related 2 (PVRL2) and hydroxysteroid dehydrogenase-like 2 (HSDL2). There is little literature regarding these genes and nodes, and it presents an exciting starting point for future studies on atherosclerotic lesion development.*
41. Kleemann R, Verschuren L, van Erk M et al.: Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis. Genome Biol. 2007, 8: R200.
42. Wheelock C, Wheelock A, Kawashima S et al.: Systems biology approaches and pathway tools for investigating cardiovascular disease. Mol Biosyst. 2009, 5: 588-602.
43. •• van Erk M, Wopereis S, Rubingh C et al.: Insight in modulation of inflammation in response to diclofenac intervention: a human intervention study. BMC Med Genomics 2010, 3: 5. *The authors examined obesity associated inflammation (cardiovascular disease related) following administration of the anti-inflammatory drug dicloflenac to overweight test subjects. This study combined results from transcriptomics, proteomics, metabolomics, oxylipin, and RNA data using multivariate analysis and Metacore v4.7.*
44. •• Inouye M, Silander K, Hamalainen E et al.: An immune response network associated with blood lipid levels. PLoS Genetics 2010, 6: e1001113. *The authors created a comprehensive study that implicates a previously uncharacterized tissue-specific gene network, the Lipid Leukocyte module, to be associated with blood lipid mediation and inflammation response.*
45. Back M, Bu DX, Branstrom R et al.: Leukotriene B4 signaling through NF-kappaB-dependent BLT1 receptors on vascular smooth muscle cells in atherosclerosis and intimal hyperplasia. Proc Natl Acad Sci U S A 2005, 102: 17501-6.
46. Yu Y, Lucitt MB, Stubbe J et al.: Prostaglandin F2alpha elevates blood pressure and promotes atherosclerosis. Proc Natl Acad Sci U S A 2009, 106: 7985-90.
47. Samuelsson B, Dahlen SE, Lindgren JA et al.: Leukotrienes and lipoxins: structures, biosynthesis, and biological effects. Science 1987, 237: 1171-6.
48. • Buczynski M, Dumlao D, Dennis E: Thematic Review Series: Proteomics. An integrated omics analysis of eicosanoid biology. J. Lipid Res. 2009, 50: 1015-1038. *The authors have prepared a detailed and thorough review pertaining to eicosanoid systems biology. It provides an excellent reference for future proteomic analysis of eicosanoids and their associated biological networks.*
49. Mitchell J, Warner T: COX isoforms in the cardiovascular system: understanding the activities of non-steroidal anti-inflammatory drugs. Nature Reviews Drug Discovery 2006, 5: 75-86.
50. Gertow K, Nobili E, Folkersen L et al.: Expression of 12- and 15-lipoxygenase mRNAs in human carotid atherosclerotic lesions: associations with cerebrovascular symptoms. Submitted to Atherosclerosis 2010.