Know Me! Unraveling the Riddle of Calcific Aortic Valve Disease by Bioinformatics

Ares Pasipoularides

Emeritus Faculty of Surgery and of Biomedical Engineering, Duke University School of Medicine and Graduate School, Durham, North Carolina, USA


Δεινόν το γήρας, ου γάρ έρχεται μόνον
Daunting is old age, for it never comes alone—
Menander, Greek poet, dramatist and comedic author (342–291 BCE)

The normal aortic valve (AV) leaflets are composed of specialized valvular endothelial and interstitial cells akin to fibroblasts and an extracellular matrix (ECM) that contains collagen, elastin and glycosaminoglycans. Rheumatic heart valve disease, erstwhile significant cause of aortic stenosis (AS), is nowadays rare in industrialized countries. Calcific aortic valve disease (CAVD) is expressed in old-age as ectopic mineralization (calcification) and fibrosis of the tri-leaflet aortic valve, and it is commonly described as age-related degenerative or dystrophic calcific AS (Pasipoularides 2010; Pasipoularides et al. 2016a, b). Some 2% of the population have congenitally bicuspid aortic valves, which are much more likely to calcify than tricuspid ones, and about one-half of these will develop CAVD. CAVD is not a binary condition with a 2-stage status: present/absent; rather it exhibits a continuum from mild to severe, and may affect to some extent as many as 25-30% of adults ≥ 65 yrs. old (Otto et al. 1999).

With disease progression, calcification and fibrosis increase leaflet stiffness and restrict orifice opening, increasing the transvalvular ejection pressure gradient and augmenting the systolic stresses that must be generated by the myocardium to overcome the obstruction (Bird et al. 1982; Pasipoularides 1990; Georgiadis et al. 1992). Inotropic reserve and the development of compensatory LV hypertrophy initially can compensate for the increased afterload (Pasipoularides et al. 1984; Shim et al. 1992; Pasipoularides 2007). However there eventually ensue worsening signs and proliferating symptoms of outflow obstruction (Fig. 1) accompanied by reduced diastolic LV compliance (Pasipoularides 1992, 2013a), subendocardial ischemia with myocardial diastolic relaxation abnormalities and cardiac arrhythmias (Mirsy and Pasipoularides 1980, 1990; Pasipoularides et al. 1985, 1986), heart failure and mortality (Pasipoularides 2013b, 2015a; Komamura et al. 1992).

It was probably Leonardo da Vinci (1452-1519) (Pasipoularides 2010, 2014) who first recognized atherosclerotic macroscopic changes. When he illustrated arterial atherosclerotic lesions in an elderly man at autopsy, he attributed the thickening of the vessel wall to “excessive nourishment” from the blood. In 1904, the eminent German physiologist-pathologist Johann Georg Mönckeberg (1877-1925), first described CAVD in Virchows Archiv, as a passive process associated with rheumatic fever or aging (Mönckeberg 1904). The recognition that CAVD arises from active cellular mechanisms involving genetic epigenetic and environmental interactions implies that the causative processes might be investigated and targeted to treat medically (Pasipoularides 2016a, b). This is the motivation for the application of a number of emerging genomic bioinformatics approaches to the analysis and characterization of key genes involved in the initiation and progression of CAVD, several of which are utilized in the trailblazing Article by Liu et al. (2017), in this issue of The Tohoku Journal of Experimental Medicine, which opens doors to CAVD not envisioned by most medical investigators just a short time ago.

Intricate interactions between genes and the environment contribute to the development of CAVD. Valve cells and the extracellular microenvironment are dynamic and reciprocally regulated; interstitial cells interact with their microenvironment directly via integrins, with biochemical signals through receptors, and with adjacent cells via cadherins (Pasipoularides 2015b, c). The valvular endothelium is an important regulator of (patho-)physiological processes, and lipids and inflammation are important mediators of atherogenesis and CAVD (Rajamannan et al 2011). CAVD had formerly been considered as a passive and non-modifiable wear-and-tear disease process that comes along with advancing age (see epigraph). However, considerable
evidence has now accumulated from molecular studies supporting a cell-mediated, active disease process involving risk factors and histopathophysiological features (endothelial cell and macrophage activation, proteolytic activity, and osteogenesis in inflamed valves) that characterize atherosclerosis (Rajamannan et al. 2011; Yutzey et al. 2014; Pasipoularides 2016a, b).

Bioinformatics Tools and Approaches for Studying the Genetics of CAVD

Deriving biological information from polygenic sets implicated in CAVD

In complex multifactorial diseases, many common genetic variants occur and have a significant frequency in the population. For such diseases genome wide association studies (GWAS) are imperative in order to identify clusters of common variants partaking in their pathogenesis. With the burgeoning of the molecular revolution and the accompanying avalanche of bioinformatic mega-databases, new comprehensive methods have been and are being developed for investigating the genetic basis of polygenic traits and diseases, including CAVD (Pasipoularides 2010). Concomitantly, modern biologic methodology has shifted to methods for genomic-scale analysis like microarray and massively parallel technologies, which allow simultaneous assessing of numerous genes in a batch (Pasipoularides 2017). This has engendered a need for tools facilitating interpretation of biological data in batch-mode, to help pluridisciplinary data examination and quantitative analyses with interpretation of the findings. Thus, we can best integrate the pieces of the massive genomic jigsaw puzzle to discern the polygenic architecture of the complex CAVD syndrome.

Microarray and next generation sequencing (NGS)
analyses allow the identification of a number of gene sets that are consistently under- or overexpressed in a phenotype/disease under genetic investigation, and present opportunities and challenges for data mining; the identified altered gene sets may contain numerous genes that are known to be interlinked in various metabolic and biological processes. The ensuing problem with the analysis of such data is that the findings at hand can only be interpreted with preexisting knowledge about gene variations that are observed in the disorder being examined; accordingly, empirical and at times experimental data must be procured for developing a knowledgeable hypothesis. In the present study, this was done by Liu et al. (2017) who used publicly available gene expression datasets of calcific and normal human aortic valve tissues, as well as a dataset containing expression profiles of in vitro fibrosal human aortic valve endothelial cells (hHVECs) that were exposed to laminar or oscillatory shear stress. Additionally, the voluminous data need to be handled skilfully to extract the clinically noteworthy facts in a clear and robust combination, using proper statistical analysis packages.

To meet data interpretation challenges, gene-set, or cluster testing is developing into an innovative analytical tool for the raw data generated from batch-mode testing, which allow us to scrutinize comprehensively the expression state of large numbers of genes in a single experiment. The raw data can only be descriptive and derived from a relatively small number of tissue samples. However, consistent alterations found in gene expression may allow a better comprehension of pathogenetic aspects and of the molecular characterization of individual disorders such as CAVD.

Weighted gene coexpression network analysis (WGCNA)

To reduce the dimensionality of the data (data-downsizing), genes and their protein products can be grouped into functional modules (clusters), according to the encoded cellular processes and pathways, and gene coexpression networks are defined so as to describe the biological relationships between the gene-transcripts involved (Pasipoularides 2015d). Each mechanism which contributes to the overall phenotype requires the collaboration of many genes. Coexpressed collaborative systems emerge via coupling of genes which together determine biological pathways associated with transcription. The term eigengene conveys the notion of a characteristic gene as representative of an entire gene co-expression module. However, the impact of individual genes on coexpressed collaborative processes is uneven. Genes that occupy core nodes of regulatory networks (hub genes) are usually tasked with apposite routing of biological signals, and highly connected hub genes in a module play important roles in normal/abnormal biological processes. Their concentrated functional involvement and interaction with complex regulatory processes may lead to raised susceptibility to mutations.

WGCNA can be used for discovering clusters of highly correlated genes, for summarizing such clusters using the module eigengene or an intramodular hub gene, for relating modules to one another and to external sample traits (using eigengene network methodology), and for calculating module membership measures. This was done by Liu et al. (2017) to explore genes and gene modules related to CAVD. Genes with expression levels that are highly correlated are biologically intriguing, since they imply common regulatory mechanisms, or participation in related biological processes. To construct a coexpression network from microarray gene-expression data, we begin by calculating essentially the Pearson correlations (Pearson’s r, perhaps the most common statistical test for determining if there is an association between entities) for all pairs of genes in the network. These correlations represent the connection strengths between genes in the network, which leads us naturally to the concept of the weighted gene co-expression network (WGCN). However, as gene expression profiles are not always correlated linearly, nonlinearly correlated genes might inadvertently be missed in the inferred co-expression network.

Gene co-expression networks (GCNs) are valuable in exploring the functionality of genes and analyzing complex gene-gene interactions, and the integration of expression data offers significant insights to understand the transcriptional mechanisms in health and disease (Allen et al. 2012). A set of genes are interacting with each, if they become more relevant when considered together; moreover, a gene could lose functionality/relevance due to the absence of any other gene interacting with it.

Genetic risk scores and liability to genetic risk

Correlation networks boost network-based gene screening techniques used to find candidate biomarkers or therapeutic targets. Gene-set interpretation methods and software packages are now accessible, which apply a variety of sophisticated statistical tests to assess the functional annotation for the genes of any specified set of interest (Subramanian et al. 2005). GWAS identify hundreds of genetic variants that are credibly associated with cardiovascular syndromes and outcomes, subclinical cardiovascular phenotypes, and risk factors for cardiovascular disease (CVD) (Pasipoularides 2015d, 2017). However, the numerous interacting individual genetic variants identified typically explain a very small fraction of the variation in complex traits such as CAVD and thus have limited predictive capacity for the resulting overall disease risk.

As the aim of genetic/genomic research is expanding in the post-Human Genome Project era, it is becoming feasible to identify genes underlying stable genetic contributions to various human traits and diseases. The genetic code is “unraveled” by gene expression, and the properties/actions of its interacting products give rise to the organism’s phenotype, encompassing normal and/or abnormal traits. Operationally, individual CVD-risk can be viewed as particular values of a continuous variable termed liability.
According to the liability-threshold model (Fig. 2), all the factors influencing the development of a multifactorial disorder, whether genetic or environmental, determine the single entity/variable known as liability, which embodies a combination of several “bad”-variant genes and environmental factors.

The liabilities of all individuals in a population establish a continuous variable, exemplified by a bell-shaped curve. Such a Gaussian curve is generated by many genes, labeled as polygenes, each acting in an additive fashion to engender the CAVD liability. To account for a discrete or binary phenotype (affected, or not) with an underlying continuous distribution, a threshold is envisaged above which the abnormal phenotype is expressed as CAVD; thus, individuals to the right of the threshold-line represent those affected by the disorder (Fig. 1). The liability of individuals actually diagnosed with the disorder is ≥ the threshold on such a liability scale. Continued investigative efforts focus on expanding our knowledge of CVD genomics (O’Donnell and Nabel 2011; Pasipoularides 2017). Aggregating information about multiple variants, each with small effects, into a single genetic risk score (GRS), or polygenic risk score, is useful for examining the cumulative predictive ability of genetic variation at known loci on the phenotype (Krarup et al. 2015). The use of polygenic risk scores in human studies is increasing (Spiliopoulou et al. 2015); their benefit is that they can be used for prediction of disease susceptibility.

**Functional annotation of genes and gene ontology**

A constructive type of testing that is nicely exemplified in the study by Liu et al (2017) involves monitoring gene expression while cells or tissues of interest undergo some biological disease process. It is increasingly clear

---

**Fig. 2.** The liability-threshold model applied to CAVD.

A. According to the liability-threshold model, which applies to CAVD, all the factors influencing the development of a multifactorial disorder, whether genetic or environmental, determine a single entity/variable known as liability that embodies a combination of several “bad” genes and environmental factors. The liabilities of all individuals in a population establish a continuous variable, exemplified by a bell-shaped curve. Such a Gaussian curve is generated by many genes, known as polygenes, each acting in an additive, cumulative fashion to engender the CAVD liability. To account for a discrete or binary phenotype (affected, or not) with an underlying continuous distribution, a threshold is envisaged above which the abnormal phenotype is expressed; thus, individuals on the right side of the threshold-line represent those affected by CAVD. In the general population, the proportion beyond the threshold defines the population incidence of CAVD, while among relatives the proportion beyond the threshold specifies the familial incidence.

B. The liability curve for relatives of affected individuals will be shifted to the right, reflecting the higher familial CAVD incidence, accountable by genomic and/or environmental (diet including ultra-processed energy-dense foods, lifestyle, drug treatment, the microbiome, etc.) factors.
that a manifest biological function can rarely be attributed to an individual gene or molecule. Instead, most normal or abnormal phenotypic traits arise from complex interactions between numerous genes and metabolomic constituents. Therefore, it is essential to understand the structure and the dynamics of the complex web of interactions that contribute to the investigated structure and function of a living cell, tissue, or organ (the aortic valve leaflets in CAVD). After gene clustering analyses and statistical selection of the genes that are significantly modulated or clustered on the basis of expression profiles, there result one or more modules or lists of highly interconnected genes that might define a biological theme, as is shown in the study by Liu et al. (2017). We can view modules as basic system constituents, i.e., nodes of a network. We then may determine whether there is a common biological theme in the particular modules or gene clusters, and may also compare the themes among the clusters. Functional annotation of genes is the process of identifying the locations of individual genes and other features on raw DNA sequences, called assemblies, implicated in a normal or abnormal phenotype and determining what those genes do—an annotation (irrespective of the context) is a note added by way of explanation or commentary. Once a genomic cluster or a WGCN is sequenced, it needs to be annotated to make sense of it. Usually genes are annotated one-at-a-time using internet-based gene ontology databases or manual literature searches, seeking to identify the most salient biological themes in order to make sense of the findings.

Founded in 1998, the Gene Ontology project (GO, a set of concepts and terms that convey gene properties and the relations between them, http://www.geneontology.org) represents a comprehensive source for functional genomics, supplying information about gene product function. It is a collaborative initiative to address the need for consistent descriptions of gene products across databases, supported by the US National Human Genome Research Institute (NHGRI) (The Gene Ontology Consortium 2015). It has developed 3 structured ontologies that designate gene products in a species-independent manner in terms of their associated biological processes, cellular components, and molecular functions. Notably, compared to the number of GO terms added to describe molecular functions and cellular components, the number of terms to describe biological processes such as are implicated in CAVD has, in recent years, increased at a higher rate.

**Integrated Bioinformatics Analysis Identifies Key Genes Involved in CAVD**

CAVD uncorrected and progressing to severe symptomatic AS is a generally fatal disease, causing anguish through angina, syncope, and heart failure before the ensuing death. Hopefully transapical and percutaneous approaches complementary to conventional surgery will make life-preserving valve replacement available to virtually all those aged people who need it. However, the promising findings and accomplishments of the pioneering study of Liu et al. (2017) in this issue of *The Tohoku Journal of Experimental Medicine*, open new gates to CAVD prevention and management. Several key genes that are involved in the development and progression of CAVD were identified, including *KLF2, PLPP3, ID1, KIT, PLAU* and *HMOX1*, which are involved in mechanos-responses; *eNOS*, which is involved in oxidative stress; *VCAM-1*, *MMPs* and *IL-8*, which are involved in inflammation; and *CTHRC1* and *SCG2*, which are involved in extracellular remodeling.

As the authors observe (Liu et al. 2017), mechanoresponse, oxidative stress, inflammation and extracellular remodeling pathways jointly form an indivisible network bringing about CAVD development (Pasioularides 2016a, b). Aberrant mechanical stress inhibits flow-sensitive transcription factors (such as KLF2), thereby initiating CAVD development. Irregular mechanical stress can lower the expression of anti-oxidation genes, such as *PLPP3* and *eNOS*, causing eNOS uncoupling and oxidative stress. Oxidative stress promotes endothelial damage by the peroxidation of lipid metabolites, activating several procalcification pathways, such as the LPA/LAPR1/NF-κB pathway that links metabolism and inflammation, damaging the integrity of the endothelium and promoting the expression of several monocyte adhesion molecules and inflammatory mediators, such as *VCAM-1* and *IL-8*. Inflammation, characterized by infiltration of leukocytes, can escalate oxidative stress and extracellular matrix remodeling, mediated by gene products, such as *MMPs*, *CTHRC1* and *SCG2*. These all promote the procalcification phenotype shift in the valvular microenvironment and can further exacerbate CAVD progression.

Metabolic disorders including atherosclerosis and CAVD have been commonly viewed as lipid storage disorders. It is now being realized that chronic low-grade inflammation partakes decisively in the initiation, propagation, and development of metabolic disorders. The findings of Liu et al. (2017) are in harmony with numerous recent studies that have implicated the transcription factor NF-κB in the development of such diseases, thus corroborating inflammation as a crucial etiologic factor and offering hope for the development of innovative approaches to CAVD prevention and treatment. Concurrently, future multidisciplinary studies will characterize the adaptive/maladaptive nature of the AVS-induced hypertrophy, its gender- and individual patient-dependent peculiarities, and its response to surgical/medical interventions (Pasioularides 2016a, b). They will herald more effective, precision medicine treatments (Pasioularides 2015d, 2017).

**Conclusions and Future Directions**

Assessing the likely cardiological applications of genome/bioinformatics projects is not easy, although enough was known before modern-day successes to suggest that the bioinformatics study of CAVD at the level of mole-
cules and cells would bestow major clinical benefits. Likewise, continued investigations into lipid receptors, cell stress pathways and the networks of inflammatory propagation should provide new and better targets for prevention and therapeutic intervention in CAVD in the future. Genetics’ clinical usefulness rests on the notion of gene causal-variants resulting in or predisposing to diseases, such as CAVD. Inheritable form-and-function variations may generally accrue from new genetic combinations through meiosis, viable flaws during cellular DNA replication, or mutations elicited by miscellaneous factors. At present, personalized genetic testing is an emergent technology. As highlighted in the approach of Liu and his collaborators (2017), we should be cautious applying it without accompanying comprehensive phenotypic information and gene ontology, which represent a comprehensive source for functional genomics and supply information about gene properties and the relations between them, as well as about gene product function. Genotype ≠ Phenotype. However, advances in genomics and bioinformatics over coming years will undoubtedly build on what has been attained and, combined with insights into how epigenetics and environmental factors affect cardiovascular health, have the potential to eradicate the misery of maladies such as CAVD and to improve human health and well-being – a tantalizing prospect.

Funding

Research support, for work from my Laboratory surveyed here, was provided by: National Heart, Lung, and Blood Institute, Grant R01 HL 050446; National Science Foundation, Grant CDR 8622201; and North Carolina Supercomputing Center and Cray Research.

References


