Journal of Proteomics 171 (2018) 37–52

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

How to use and integrate bioinformatics tools to compare proteomic data from distinct conditions? A tutorial using the pathological similarities between Aortic Valve Stenosis and Coronary Artery Disease as a case-study

Fábio Trindade [a,b,*], Rita Ferreira [c], Beatriz Magalhães [a], Adelino Leite-Moreira [b], Inês Falcão-Pires [b], Rui Vitorino [a,b]

[a] Institute of Biomedicine, Department of Medical Sciences, University of Aveiro, Aveiro, Portugal
[b] Unidade de Investigação Cardiovascular, Departamento de Cirurgia e Fisiologia, Faculdade de Medicina, Universidade do Porto, Porto, Portugal
[c] QOPNA, Mass Spectrometry Center, Department of Chemistry, University of Aveiro, Aveiro, Portugal

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 28 February 2017
Accepted 19 March 2017
Available online 21 March 2017

Keywords:
Proteomics
Bioinformatics
STRING
DisGeNET
Cytoscape
ClueGO

ABSTRACT

Nowadays we are surrounded by a plethora of bioinformatics tools, powerful enough to deal with the large amounts of data arising from proteomic studies, but whose application is sometimes hard to find. Therefore, we used a specific clinical problem – to discriminate pathophysiology and potential biomarkers between two similar cardiovascular diseases, aortic valve stenosis (AVS) and coronary artery disease (CAD) – to make a step-by-step guide through four bioinformatics tools: STRING, DisGeNET, Cytoscape and ClueGO. Proteome data was collected from articles available on PubMed centered on proteomic studies enrolling subjects with AVS or CAD. Through the analysis of gene ontology provided by STRING and ClueGO we could find specific biological phenomena associated with AVS, such as down-regulation of elastic fiber assembly, and with CAD, such as up-regulation of plasminogen activation. Moreover, through Cytoscape and DisGeNET we could pinpoint surrogate markers either for AVS (e.g. popeye domain containing protein 2 and 28S ribosomal protein S36, mitochondrial) or for CAD (e.g. ankyrin repeat and SOCS box protein 7) which deserve future validation. Data recycling and integration as well as research orientation are among the main advantages of resorting to bioinformatics analysis, hence these tutorials can be of great convenience for proteomics investigators.

Biological significance: As we saw for aortic valve stenosis and coronary artery disease, it can be of great relevance to perform preliminary bioinformatics analysis with already published proteomics data. It not only saves us time in the lab (avoiding work duplication) as it points out new hypothesis to explain the phenotypical presentation of the diseases as well as new surrogate markers with clinical relevance, deserving future scrutiny. These essential steps can be easily overcome if one follows the steps proposed in our tutorial for STRING, DisGeNET, Cytoscape and ClueGO utilization.

1. Introduction

Large volumes of data are often obtained after the application of high throughput technologies, such as mass spectrometry, for the analysis of biological samples in the advent of disease proteomics. Commonly, the utilization of these highly sensitive technologies is chosen when trying to answer a question formulated a priori or to get the "big picture" of the biological processes underlying the pathogenesis of a given condition. Thus, enormous amounts of data are left to analyze and can be object of a second analysis by widely available bioinformatics tools. Such data recycling can be important to answer to certain biological/etiological questions or even to guide research towards the response of specific questions meanwhile raised during bioinformatics analysis. To help with that task we have a panoply of web-tools and programs at disposal that help integrate protein datasets and extract biological knowledge from their predicted or already proved interactions. These include the web-available STRING v10 (http://string-db.org/, [1]) and DisGeNET (http://disgenet.org/web/DisGeNET/menu, [2]) as well as the software applications Cytoscape (http://www.cytoscape.org/, [3]) and its plug-in ClueGO [4], to name a few.

STRING provides a network view on functional protein associations, based on direct (physical) and indirect (functional) protein-protein interactions [1]. This webtool uses both known and predicted interactions (derived from indirect evidences of gene co-occurrence, fusion events, co-expression and conserved neighborhood), to whom a confidence

* Corresponding author at: iBiMED, Institute of Biomedicine, Department of Medical Sciences, University of Aveiro, 3810-193 Aveiro, Portugal.
E-mail address: fabiotrindade@ua.pt (F. Trindade).

http://dx.doi.org/10.1016/j.jprot.2017.03.015
1874-3919/

score is attributed by comparing to a reference set of trusted true associations (KEGG database). In another words, the protein-protein interaction score is the probability of the existence of such interaction in KEGG database [5]. This tool can be of utility when one seeks to take insight into the biological processes that might be involved in the pathogenesis of diseases, because it provides the gene ontology (GO) annotation of the inputted proteins. Furthermore, STRING can be important to look for surrogate markers of diseases because it allows to determine proteins that interact with the inputted dataset (and, thus, that may be involved in their pathogenesis).

DisGeNET, in turn, is a comprehensive, expert-curated repository of gene-disease associations (GDA) [2]. Even though it entails associations of non-coding genes, e.g. microRNA, with diseases, we can easily work with proteins, as they are presented with their UniProt KB accession code. The major advantage of using DisGeNET database is the opportunity to bypass literature search and data-mining and with a few clicks we can check which genes/proteins (from now on proteins) are already associated with a disease (please see step 3 of the tutorial) and if they are disease-specific or, on the contrary, if they have been linked with more than one condition (please see step 5 of the tutorial). Moreover, DisGeNET provides a score and a disease specificity index along with stored associations, providing the user with a measure of the strength of that association and the specificity of the gene/protein in question, respectively. Such score takes into consideration the number of sources where the deemed associations were pointed, the kind of data curation, the type of animal models used as well as the number of publications derived from text-mining sources [2].

Although STRING provides a fast way to gather the potentially deregulated biological processes for a given disease, the Cytoscape's apps ClueGO and CluePedia deliver, in a more integrative and dynamic way, which of those are markedly differentiated across two or more conditions. This is due to the ability to perform cluster-based analysis [4]. Probably, the time-consuming and memory-demanding nature of these programs are the only limitations comparing to the aforementioned web-tools. Actually, STRING and DisGeNET were designed with a Cytoscape plugin, showing how powerful this program can be. It should be noted that CluePedia can be used independently of ClueGO, but we find that using both applications provides the best output in terms of GO enrichment analysis. When used together, these applications allow to scrutinize specific biological processes to each condition and, simultaneously, to detect which proteins are involved in those phenomena. Another relevant information we can take from these tools is the deregulated biological processes in AVS and CAD. That can be accomplished using the Cytoscape's apps ClueGO and CluePedia which perform GO enrichment analysis and include the information associated to the biological processes, molecular function or pathways into the network. Such network is created first by mapping associated genes to the set ontologies (nodes), which are then connected based on shared genes (kappa score – a measure of the association strength).

Herein, we provide an example of how revisiting proteomic data through database and gene ontology (GO) analysis with these tools can help us to extract biological knowledge from already published proteomics papers. It is not our intention to compile users manuals/tutorials of these tools or databases. Instead, we propose a protocol to extract the most relevant biological information out of these resources and show how they can be useful before setting new lines of research. In order to demonstrate this workflow, we will take a clinical problem – how to discern the pathophysiology of two related pathologies [aortic valve stenosis (AVS) and coronary artery disease (CAD)] and which proteins can be proposed as surrogate markers – as a step-by-step example to demonstrate their applicability. Thus, we will start by giving a brief summary of the pathogenesis of both diseases and then we will look to proteomic data collected from both disease settings and analyze through STRING, DisGeNET, Cytoscape and its ClueGO app, giving the reader a tutorial perspective over these tools. In Supplementary Material, the reader can find a detailed discussion of the biological phenomena underlying AVS and CAD pathogenesis, where some biological questions deserving future scrutiny are raised.

2. Methods

2.1. Literature search

In order to collect data available from proteomic studies in AVS and CAD settings, independent PubMed searches were ensued up to 19th July 2016, by two different users using the following keywords in separate queries:

- "aortic valve stenosis proteomics",
- "aortic valve disease proteomics",
- "aortic valve stenosis proteome",
- "aortic valve disease proteome",
- "coronary artery disease proteomics",
- "ischemic heart disease proteomics",
- "coronary artery disease proteome" and
- "ischemic heart disease proteome".

Two experienced, independent reviewers pre-selected a list of articles potentially relevant to extract data for further bioinformatics analysis. Any disagreement was subsequently resolved between the two. Only full-size English-written articles published in peer-reviewed journals were considered, taking into consideration the following inclusion criteria:

- proteomic studies,
- studies enrolling humans only,
- studies solely enrolling subjects with AVS (or CAD) in comparison to healthy individuals or without any known cardiovascular disease, except for CAD (or AVS) and
- studies enrolling samples such as plasma, serum, urine, aortic valve, coronary arteries, valve-derived and coronary-arteries derived cells and tissues (except for [6] which compared directly AVS and CAD in myocardial biopsies).

As an exclusion criterion, all studies related to pharmacological or interventional studies were left apart. Then, selections by both users were crossed and a final consensus was reached with regard to the studies to include in the analysis.

2.2. Data mining

Data from each study was extracted to Excel spreadsheets and the following fields were filled in: "Protein ID" (UniProt code), "Protein Name", "Variation" (+, −, unchanged or N/A – not applicable), "fold-change" (if available), "Sample" (biofluid, biopsy, secretome), "Specific Condition" (concrete pathological setting – AVS or CAD), "Sample Size N (n validation)" (number of enrolled patients for discovery – N – and validation – n – phases), "Background Condition" (pathophysiological status of the subjects without the condition of interest), "Sample Size N (n validation)" (number of enrolled subjects, without AVS or CAD, for discovery – N – and validation – n – phases), "Experimental approach", "Validation" (experimental approach used for validation), "fold-change" (if available), "*p*-value" (if available), "reference", "PMID" (PubMED article reference number) and "Source of Data" (paper tables, figures, core text or supplementary materials). These data can be found in Supplementary Files 1 (AVS) and 2 (CAD).

In order to perform STRING, Cytoscape and ClueGO analyses, the list of proteins associated to AVS and CAD were filtered by selecting only "+" and "−" in the "Variation" field. Only consistent protein variations were selected. Protein levels were considered "unchanged" if the *p*-value was found higher than 0.05. In such cases, proteins were not included in downstream analysis.

2.3. Bioinformatics analysis

Bioinformatic analysis was ensued on STRING v10.0, DisGeNET v4.0, Cytoscape v.3.4.0, ClueGO v.2.2.6 and CluePedia v.1.2.6. following the setup described in detail in Section 4. System requirements can be found in Supplementary Materials.

3. Overview on the pathogenesis of aortic valve stenosis and coronary artery disease: Where they meet and where they diverge?

Aortic Valve Stenosis is the most common valvular disease in the western world and is associated with large economic expenses in healthcare systems. It is characterized by the progressive narrowing of the aortic valve with a consequent increase in the pressure afterload on the left ventricle, triggering a cardiac hypertrophic response firstly aiming at normalizing wall stress but sooner or later becoming maladaptive [7]. AVS represents the commonest cause of valve replacement in the developed countries [8] and, over the last decade, the concept of AVS as a degenerative disease was replaced by evidence that aortic valve calcification might represent an active inflammatory and atherosclerotic process [9]. In turn, atherosclerosis represents a hallmark of coronary artery disease (CAD), also known as ischemic heart disease. CAD remains the leading cause of death worldwide and also imposes a major burden in the public health systems [10,11]. Briefly, in either AVS or CAD, an initial endothelial lesion resulting from increased mechanical stress takes place. Then, lipid deposition (mainly low-density lipoproteins) and oxidation ensues, creating a very inflammatory and cytotoxic environment. Consequently, endothelial cells increase the expression of adhesion molecules driving the infiltration of monocytes and T cells. Inside the stenotic valves/atheroma, monocytes differentiate into macrophages and T cells release several pro-inflammatory cytokines. The paracrine cues of such cells communicate with endothelial and smooth muscle cells, in CAD, and with fibroblast-like valve interstitial cells, in AVS, which together are responsible for the secretion of matrix metalloproteases and tissue inhibitors of metalloproteases. Such molecules disorganize extracellular matrix and in combination with increased secretion of collagen by fibroblasts are responsible for increased valve or artery stiffness, in AVS and CAD, respectively. Angiogenesis is also observed in both conditions [7,8,12–17].

Despite the aforementioned similarities, AVS and CAD display also specific pathogenic features, starting in rheology. While the coronary artery is subjected to sustained laminar blood flow, the aortic valve is exposed to pulsatile shear stress on the ventricular side and low and reciprocating shear stress on the aortic side. This is thought to induce the release of transforming growth factor β1 (TGF-β1) from platelets and to activate it. TGF-β1 may, in turn, lead to valve narrowing and fibrosis, thus further increasing shear stress in a vicious cycle-manner [18]. Other differences are found in the cellular entities involved and in secondary adverse events. For instance, foam cells are more characteristic of CAD and osteoblasts are more specific of AVS (although they can be found in both cases). The former result from the transformation of macrophages when they are no longer able to mobilize cholesterol from internalized oxidized LDL. Eventually, the phagocytic activity of these cells is hampered and they undergo apoptosis leading to passive calcification [11]. The latter is thought to happen as a result of myofibroblast-like valve interstitial cells differentiation due to the activation of several osteogenesis-related pathways, which explains AVS progression to be often accompanied by active calcification and bone formation [7]. Moreover, while adverse events in CAD encompasses thrombosis due to plaque rupture or ischemic insults due to artery narrowing, in AVS one often observes progressive valve rigidity and ultimate decompensation of myocardial response [7,11,15].

Even though early lesions in AVS and CAD share several pathophysiological features, namely the process of atherosclerosis, the commonly used statin therapy for CAD, aiming to inhibit endogenous cholesterol synthesis, failed to show any impact on AVS progression [7,13,16]. This fact reflects our incomplete knowledge about its pathogenesis and also demands the investigation of novel therapeutic targets in order to avoid aortic valve replacement surgery. In the next section, we will use this problem to show how several bioinformatics tools can help us guide research to better understand the etiopathogenic routes of AVS and CAD and also towards the definition of candidate disease-specific markers. Empowered with such tools, we aim to answer the following biological questions:

- Is there any or some deregulated biological processes able to distinguish the pathogenesis or the phenotypical presentation of AVS and CAD?
- Do bone formation and metabolism-related biological processes have potential to distinguish both diseases?
- Is there any protein or group of proteins that may become surrogate markers for anticipated diagnosis, prognosis or for therapeutic monitoring?

4. Revisiting proteomic data through bioinformatics analysis

There is still a low number of studies following a proteomics approach to study the phenotypical changes in AVS and CAD and, to the best of our knowledge, only one study directly addresses the differences on the protein level among the two conditions [6]. Nevertheless, such study was carried out with ventricle biopsies, thus its data mainly reflects myocardial secondary complications of AVS and CAD (ventricular remodeling) and is not primarily centered in distinguishing the etiological routes of both pathologies. Therefore, we will take this problem as a step-by-step example to take a tutorial tour across some bioinformatics tools (STRING, DisGeNET, Cytoscape and ClueGO) and to explain how they can be useful to pinpoint potentially deregulated biological processes as well as to detect surrogate markers for the differential diagnosis of these or any other pathologies. In order to do that, we retrieved the proteome data from previous studies with regard to AVS and CAD, which are summarized in Tables 1–3. The main flowchart to perform the bioinformatics analysis is given below. The reader should be aware that the protocol is a suggestion and it is not mandatory to follow the order proposed herein, but a unifying strategy was developed to help extract the maximum information of these tools and to gather a general view over the main biological processes and protein players in these or other diseases. When performing these steps, it is useful to have the "Tutorial slides" at hand. There, the reader will find several screenshots and explanations that clarify the steps in each analysis.

1. Retrieve the proper protein identifiers. Most tools readily recognize UniProt KB identifiers.
2. Predict main protein-associated biological processes and pathways in each condition, through STRING analysis.
3. Retrieve the main disease-protein associations, that is, proteins with the highest biomarker value known to date, for each condition with DisGeNET.
4. Identify proteins specific for each condition through cluster-based Cytoscape analysis.
5. Analyze the biomarker potential of the proteins found in step 4, by searching for associated diseases in DisGeNET.
6. Evaluate deregulated biological processes with ClueGO and CluePedia in order to attribute a biological meaning.

The detailed procedures and explanations for each step is given below:

1. While organizing proteome data in spreadsheets, retrieve the UniProt KB accession code for the proteins identified by other codes. To make the conversion, just follow the next steps:
   1.1. Go to UniProt KB website (http://www.uniprot.org/) and select the "Retrieve ID/mapping" tool.
   1.2. Paste your gene/protein list in the field entitled "1. Provide your identifiers".

Table 1
Characterization of the Proteomic Studies related to Aortic Valve Stenosis enrolled in the bioinformatics analysis.

Disease subtype	Sample	Controls	Experimental approach	Number of proteins	Ref.
Degenerative AVS	Aortic Valve	Healthy valves collected in autopsy (non-related deaths)	2-DE-MALDI-TOF/TOF	25	[25]
Degenerative AVS	Aortic Valve Tissue Culture Secretome	Healthy valves collected in autopsy (non-related deaths)	SDS-PAGE-nLC-MS/MS Selected Reaction Monitoring	50	[26]
	Plasma	Subjects without known CVD	Selected Reaction Monitoring	3	
Calcific Aortic Stenosis	Serum	Healthy subjects	iTRAQ + nLC-MALDI-TOF/TOF Western Blot	169 3	[27]
Degenerative AVS	Aortic Valve	Healthy valves collected in autopsy (non-related deaths)	2D-DIGE-MALDI-TOF/TOF Western Blot IHC	17 12 5	[16]
AVS	Left Ventricular Biopsy Right Ventricular Biopsy	Coronary Artery Disease	TMT + nLC-MS/MS	9 73	[6]
Degenerative AVS	Aortic Valve	None (exploratory)	2D-DIGE-MALDI-TOF/TOF iTRAQ 2D-LC-MS/MS	15 53	[28]
Degenerative AVS	Aortic Valve	Healthy valves collected in autopsy (non-related deaths)	Selected Reaction Monitoring Western Blot	2 2	[21]
Severe Degenerative AVS	Plasma	Subjects without known CVD	2D-DIGE-MALDI-TOF/TOF Selected Reaction Monitoring Western Blot	38 6 8	[29]
AVS	Aortic Valve Leaflet (thickened and calcified area)	Aortic Valve Leaflet (non-thickened and non-calcified area)	2-DE-MALDI-TOF/TOF Western Blot	7 1	[30]
AVS	Aortic Valve (calcified tissue)	Aortic Valve (non-calcified tissue)	MALDI-IMS IHC	2 2	[31]
AVS	Aortic Valve (calcified tissue)	Aortic Valve (non-calcified tissue)	iTRAQ + nLC-MALDI-TOF/TOF Western Blot	61 6	[32]

Abbreviations: 2-DE: 2-Dimensional Electrophoresis; AVS: aortic valve stenosis; CVD: cardiovascular disease; DIGE: quantitative differential electrophoresis; IHC: immunohistochemistry; IMS: imaging MS; iTRAQ: isobaric tag for relative and absolute quantitation; LC: liquid chromatography; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; nLC: nano-LC; SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TMT: tandem mass tags; TOF: time-of-flight.

1.3. Choose the starting identifier (GenInfo (gi) number and gene name are the commonest, apart from the UniProt KB identifier) and the desired identifier (UniProt KB) in the field entitled "2. Select options".
1.4. In organism box, select "*Homo sapiens* [9606]".
1.5. Download the new list as "Target List".

An example of protein identifier conversion is given in Supplementary Table 1. The proteins were identified by Martín-Rojas et al. [16] with the accession code (in UniProt KB it is classified as UniProtKB AC/ID) and we used the "Retrieve ID/mapping tool" (UniProtKB/Swiss-Prot UniProt release 2016_11) to get the UniProt KB accession. The latter is compatible with all tools discussed in this tutorial.

Note: Unrecognized identifiers should be manually curated using UniProtKB query main box. Any dubious association should be dismissed and not further considered in downstream bioinformatics analysis.

Example 1. Poduri et al. [19] identified glutathione transferase in their experiment, but they have not indicated its identifier. With the protein name only, we cannot accurately discriminate which protein they have identified. UniProt KB returns 46 reviewed human proteins with the terms "glutathione transferase". Thus, we have to dismiss this protein for further analysis.

Example 2. Identification of transforming growth factor beta by Lee et al. [20]. The authors report this protein with its gi identifier (gi|215,794,746). However, UniProt KB does not recognize such entry. To bypass this issue, we have to search by protein name in the query box. We can attribute the UniProt KB ID to this protein (P01137) because this is the only protein with such name.

Example 3. Mapping of the up-regulated proteins in AVS reported by Martin-Rojas et al. [21]. When one uploads the complete set of proteins (identified by gi number) in the UniProt KB's mapping tool we can see that 33 out of the 37 imported proteins were mapped to 57 UniProt KB IDs. This happens because the same gi number can be mapped to different UniProt KB IDs. Usually, filtering the results to depict only reviewed entries gives the exact correspondences. Always check for protein name correspondence. If necessary, check the alternative names by clicking on the entry link (marked in blue). This is important because sometimes authors give non-recommended names to the proteins. After mapping, missing proteins can be addressed by searching individually in the query box. In this case, we could confirm individually periostin and orosumucoid 1 that were initially unmapped.

2. In order to predict which biological processes and pathways are related to the proteins associated to each condition, we can perform a STRING analysis:
2.1. In STRING website (http://string-db.org/) perform a search for "Multiple Proteins"
2.1.1. Open the "Multiple Proteins" tab and paste the list of proteins in the field "List Of Names:". Alternatively, you may upload a .txt file.

Note: The user may use data found in Supplementary File 3 to perform this analysis.

2.1.2. Choose "*Homo sapiens*" as the organism of interest.
2.1.3. Check the associations of the input list with those found by STRING. When all associations are correct press "continue".

Table 2
Characterization of the Proteomic Studies related to Coronary Artery Disease enrolled in the bioinformatics analysis.

Disease subtype	Sample	Controls	Experimental approach	Number of proteins	Ref.
CAD	Plasma HDL	Subjects with no evidence of CAD	LC-MS	15	[33]
CAD	Urine	Subjects with no evidence of CAD	1-DE-nLC-ESI MS/MS ELISA	14 1	[20]
CAD	Plasma	Subjects without known CVD	iTRAQ + SCX-RPLC-MALDI-TOF/TOF Biochemical Analyser ELISA	21 2 7	[34]
CAD	Plasma HDL	Healthy subjects	iTRAQ + 2D nLC-MS/MS ELISA	12 2	[35]
CAD	Serum HDL	Healthy subjects	nLC-ESI-MS/MS ELISA	74 2	[36]
CAD	Epicardial Adipose Tissue Secretome	Subjects without known CVD	1-DE-MS ELISA	3 3	[37]
CAD	Plasma	Subjects with no evidence of CAD	SDS-PAGE-CX-RPLC-ESI-MS/MS	23	[38]
CAD	Plasma	Subjects with no evidence of CAD	SELDI-TOF-MS	32	[39]
CAD	Blood monocytes	Healthy subjects	2-DE MALDI-MS	5	[19]
CAD	Urine	Healthy subjects	CE-ESI-TOF-MS	8	[40]
CAD	Plasma Granulocytes	Subjects with no evidence of CAD	Label Free Quantification + LC-MS/MS	27	[41]
CAD	Urine	Subjects with no evidence of CAD	CE-ESI-TOF-MS IHC	2 2	[42]
CAD (with stable angina)	Platelets	Subjects with no evidence of CAD	2-DE-ESI-MS/MS (Ion-trap) Spectrophotometric Enzymatic Assay Western Blot	4 1 3	[43]
CAD	Plasma HDL_3	Healthy subjects	SCX-RPLC-ESI-MS/MS ELISA	5 1	[44]
CAD	Coronary Arteries	Normal arteries (non-stenotic)	2-DE-LC-MS/MS (Ion Trap) Western Blot	1 1	[45]
CAD	Plasma HDL_2	Healthy subjects	LC-MALDI-TOF/TOF	3	[46]
CAD	Atherosclerotic Coronary Biopsy or Necropsy	Preatheroscle-rotic Radial Biopsy or Necropsy	2D–DIGE-MALDI-TOF/TOF IHC	12 5	[47]
CAD	Plasma	Subjects with no evidence of CAD	Western Blot	1	[48]
CAD	Serum and Peripheral Blood Mononuclear Cells	Healthy subjects	2-DE-MALDI-TOF/TOF Western Blot ELISA	25 1 1	[24]
CAD	Plasma	Healthy subjects	Automatic Measurement Sandwich Immunometric Assay ELISA	3 1 4	[49]

Abbreviations: 2-DE: 2-Dimensional Electrophoresis; CAD: coronary artery disease; CVD: cardiovascular disease; CX: cation exchange; DIGE: quantitative differential electrophoresis; ELISA: enzyme-linked immunosorbent assay; ESI: electrospray ionization; HDL: high-density lipoprotein; IHC: immunohistochemistry; iTRAQ: isobaric tag for relative and absolute quantitation; LC: liquid chromatography; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; nLC: nano-LC; RPLC: reverse phase LC; SCX: strong CX; SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SELDI: surface-enhanced laser desorption ionization; TOF: time-of-flight.

For example, when inputting CAD proteome data in STRING's query we found three conflicting associations:

- Apolipoprotein C-IV (P55056) was incorrectly associated to apolipoprotein C-II (identified as ENSG00000224916). In this case, the correct option was "APOC4 - apolipoprotein C-IV …";
- Apolipoprotein C-II (P02655) was only associated to ensemble code (ENSG00000224916). In this case, an association to the second option ("APOC2 - apolipoprotein C-II …") was missing;
- Hemoglobin alpha chain (P69905) was only linked to isoform 2 ("HBA2 - hemoglobin, alpha 2"), thus the association to isoform 1 ("HBA1 – hemoglobin, alpha 1 …") was also missing.

Table 3
Review board: deregulated proteins in aortic valve stenosis and coronary artery disease.

	Aortic valve stenosis	Coronary artery disease
# Unchanged or non-quantified proteins	209	65
# De-regulated proteins	214	185
Conflicting variations	31	16
Up-regulated	130 (out of 161)	80 (out of 96)
Down-regulated	53 (out of 84)	89 (out of 106)

2.2. After data processing you will observe a protein-protein interaction network, in which nodes represent proteins and edges their predicted or known associations. Some parameters can be adjusted in the network, but only the most relevant are indicated:

2.2.1. Data Settings. It is possible to choose which kind of interaction sources are used by STRING (e.g. text-mining, experiments and databases), as well as the minimum score for the interaction and the maximum number of interactors allowed to show.

Note: For a first analysis, it can be useful to be the least stringent possible with regard to the choice of interaction sources (that is, you should select all). Depending on the amount of proteins involved in the analysis you should decide the score of interaction and the maximum number of interactors. Larger datasets demand for lower number of interactors and vice-versa, otherwise you would get protein-crowded, undecipherable networks. Moreover, irrespective of the determined parameters, data comparison from two pathophysiological conditions is made valid only using the same criteria.

2.2.2. View Settings. The most important settings are related to the meaning of the network edges. The user has 3 options of visualization that are for different purposes. By "evidence" you can see what kind and how many different

data sources helped to build the network (of course, conditioned by the sort of interactions selected in 2.2.1.). By "confidence" you may see edges with different thicknesses according to the strength of the data collected. Finally, by "molecular action" you can see what kind of interaction proteins actually stablish, for instance green means 'activation' and red represent 'inhibition'.

2.3. Go to "Analysis" tab. In this section you will find the main network stats and the most relevant annotated biological processes, molecular functions and cellular components. To export all biological processes associated with the protein input, simply download the .tsv file on "Save/Export".
2.4. Check for surrogate markers not present in your dataset.

We used STRING essentially to have a quick look over main biological processes and pathways associated to AVS and CAD. Nonetheless, STRING can also be important in the discovery of surrogate markers for these conditions. To do that:

2.4.1. Go to "Data Settings"
2.4.2. Enrich the network with second shell interactors (proteins). To do that you should:
2.4.2.1. Choose "no >5 interactors" in the "max number of interactors to show"
2.4.2.2. "Update settings". Added proteins will appear in gray nodes, as they were not present in the original dataset.

With such interaction-based enrichment analysis we found, for instance, pyruvate dehydrogenase protein X component, mitochondrial (PDHX, UniProt code O00330) and NADH dehydrogenase (ubiquinone) iron-sulfur protein 3, mitochondrial (NDUFS3, UniProt code O75489) to be potentially associated to AVS and CAD, respectively.

3. In order to retrieve the proteins already linked to AVS and CAD and to determine their specificity to the disease at scope, we can resort to DisGeNET database:
3.1. In DisGeNET website (http://disgenet.org/web/DisGeNET/menu), go to "Search" tab.
3.2. Keep the default query on "diseases".
3.3. Type the MeSH term for the disease and perform search. You should be as general as possible, unless a specific subtype or presentation of the disease is at scope.

For instance, when the goal is to search for protein associations to aortic valve stenosis, if one types only "aortic stenosis" the dropdown list exhibits several presentations of AVS, such as "supravalvular aortic stenosis" (C0003499), "aortic stenosis symptomatic" (C0741183) or "congenital supravalvular aortic stenosis" (C1305147). Thus, to retrieve the list of associated proteins one should specifically type "aortic valve stenosis" which is the broader MeSH term to whom only one option is available: C0003507.

3.4. Below the general classifications and codes for the disease, the user is presented with the "Top 10 gene associations for this disease". For the complete list you must "Browse details…".
3.5. Check the Disease Specificity Index (DSI). This is a measure of the specificity of a particular gene/protein. It spans from 0 (meaning that it is associated with many phenotypes) to 1 (meaning it is associated to one phenotype). Ideally, we would have at least one protein with a DSI = 1 for each condition. Moreover, proteins with DSI = 1 should have high scores of association, otherwise their association need to be validated.
3.6. Export data by pressing on "Download". You may either download it as a "Tabulated text file" or as an Excel file.
3.7. In Excel, filter the results by excluding empty cellsin the "Uniprot" column. This way you will gather all known proteins associated with the pathology of interest, excluding miRNAs or non-coding genes.

4. Cytoscape program can be used to identify which proteins are associated exclusively with AVS or CAD and which of those are associated with both conditions:
4.1. Create an Excel file with protein associations to disease. There should be three columns:
- "Proteins" should list proteins retrieved by literature analysis (UniProt code);
- "Specific Condition" should indicate the respective associated diseases;
- "Background Condition" should describe the health status of the subjects (e.g. healthy subjects, non-related conditions) to whom the proteome data from the "Specific Condition" was compared.

Note: The user may use data found in Supplementary File 5 to perform this analysis.
Only manually curated up- and down-regulated proteins should be considered from all data collected.

4.2. Open Cytoscape program.
4.3. Import New Network from the Excel file created in step 1.

4.3.1. Go to "File" » "Import" » "Network" » "File" and choose the desired file. Such file should be organized in three columns, representing "source nodes", "target nodes" and "edges". In this case, UniProt accessions will be representing proteins and will be defined as the "source nodes", the specific conditions will be defined as "target nodes" and the background conditions (healthy subjects or without directly related pathologies) will be the used as "edges".
4.3.2. On the top of each column define the type of interactor. The first column ("Proteins") must be defined as the "source node" (green circle). The second column ("Specific Condition") must be defined as the "target node" (double orange circle). Finally, the third column ("Background Condition") must be set as the "interaction type" (purple triangle).
4.3.3. On "Advance Options" make sure to check if the column names are attributed to the first row. Press "OK" and you will get the network on the visualization window. Since we are comparing two pathologies, one can observe two main clusters. The network can be as difficult to read as many conditions are studied, resulting in more network clusters.

4.4. Identify which nodes are exclusively related to each disease and those who are related to both diseases by dragging first neighbors of both disease nodes to the same empty space (Fig. 1). Isolated nodes, i.e. stablishing a connection with either AVS or CAD, have a higher biomarker potential a priori.

Note: When working with three or more conditions, it is useful to use "Network Analyzer" tool. When that is the case, you should select the option "Treat the network as undirected", because we are not necessarily dealing with biological cascades or intracellular pathways. This tool is helpful to identify proteins that have multiple disease associations (thus, with lower biomarker potential and vice-versa), because it is possible to generate nodes whose size is proportional to the number of edges/interactions. Further information of this tool can be found on the tutorial: http://manual.cytoscape.org/en/stable/Network_Analyzer.

html. Fig. 2 shows the same network as in Fig. 1, after analysis with Network Analyzer. Note that protein nodes related to both diseases are now bigger, representing those with lower marker potential.

4.5. Export network.
  4.5.1. Go to "File" » "Export" » "Network View as Graphics".
  4.5.2. Select the desired file format, name it and find the proper location. Files should be preferably saved as .png.
  4.5.3. Set resolution as 600 dpi and the zoom to 500%. This way all labels will be clear.
  4.5.4. Press "OK". (Fig. 2)

4.6. Save session. At any point, one may save the current session. It is advisable to do this throughout the analysis in order to avoid the loss of already executed analysis.
  4.6.1. Go to "File" » "Save as".
  4.6.2. Name the file and choose its location.
  4.6.3. Save the file as .cys.

5. Now, knowing which proteins are uniquely associated to either AVS or CAD, we can check if those were already associated to other diseases. For that purpose, let's go back to DisGeNET site. This approach allows us to rapidly identify new potential biomarkers and discard those proteins that are deregulated in different pathological settings.
  5.1. In DisGeNET website (http://disgenet.org/web/DisGeNET/menu), go to "Search" tab.
  5.2. Change the default query from "diseases" to "genes".
  5.3. Type the protein's name, UniProt identifier or its gene symbol (available on NCBI: https://www.ncbi.nlm.nih.gov/gene/)

For instance, if one is interested in searching diseases associated to vimentin you can type "vimentin", "P08670" (UniProt ID) or "VIM" (gene symbol).

5.4. Below the general information about the gene/protein, the user is presented with the "Top 10 disease associations for this gene". For the complete list you must "Browse details…".

5.5. Export data by pressing on "Download". You may either download it as a "Tabulated text file" or as an Excel file.

6. In order to identify the specific biological processes that are up- or down-regulated in AVS and CAD, through ClueGO + CluePedia one should:
  6.1. Prepare the lists of up-regulated and down-regulated proteins in both diseases, using the proper UniProt identifier.

Note: To fulfill this task, one can use scripts written in R or SQL, online Venn diagram programs such as Jvenn (http://bioinfo.genotoul.fr/jvenn/) or even Excel conditional formatting functions. We will briefly explain how to obtain the former list with the last two tools.

  6.1.1. Using Venn diagrams. For each condition, copy the proteins (UniProt KB code) that are up- and down-regulated and paste in each Venn field (List 1 and List 2). The program will look for proteins that are common to both lists and collate them in the central section of the diagram. These proteins display conflicting variations in each disease and, thus, should be discarded. To retrieve the proteins found to be uniquely up- or down-regulated simply select the non-overlapped areas on the diagram and a list is generated below Venn.
  6.1.2. Using Excel Conditional Formatting option. For each pathology, copy the proteins (UniProt KB code) that are up- and down-regulated and paste in separate columns (UP) and (DOWN). Remove duplicates in each column. Then select both columns and go to "Conditional Formatting" » "Highlight Cell Rules" and "Duplicate Values". Finally, separately filter columns to get only non-marked cells (these represent coherent up or down-regulated proteins).

As one can see in Table 3 we found 31 and 16 proteins with conflicting variations in AVS and CAD, respectively. Therefore, such proteins should be discarded.

Note: The user may use data found in Supplementary File 3 to perform the analysis from now forth. 4 clusters were defined containing

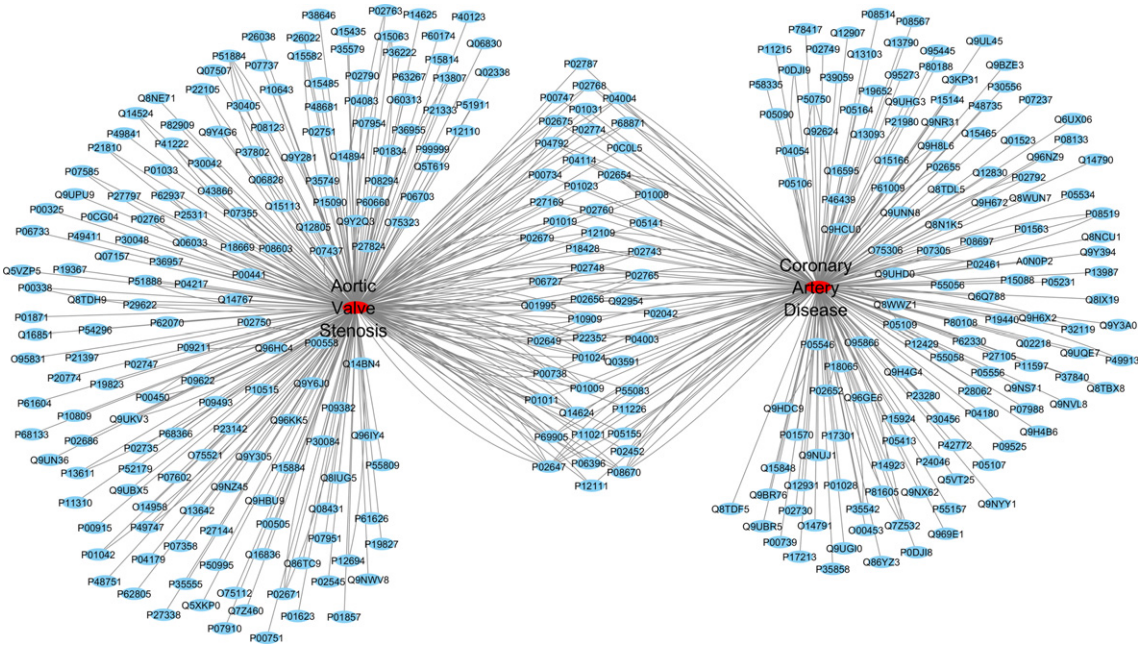

Fig. 1. Cytoscape network depicting protein associations to aortic valve stenosis and coronary artery disease (red large central nodes). Scattered blue nodes represent proteins already associated to the diseases. Proteins are represented with the respective UniProt code. The respective protein name can be found in Supplementary File 4.

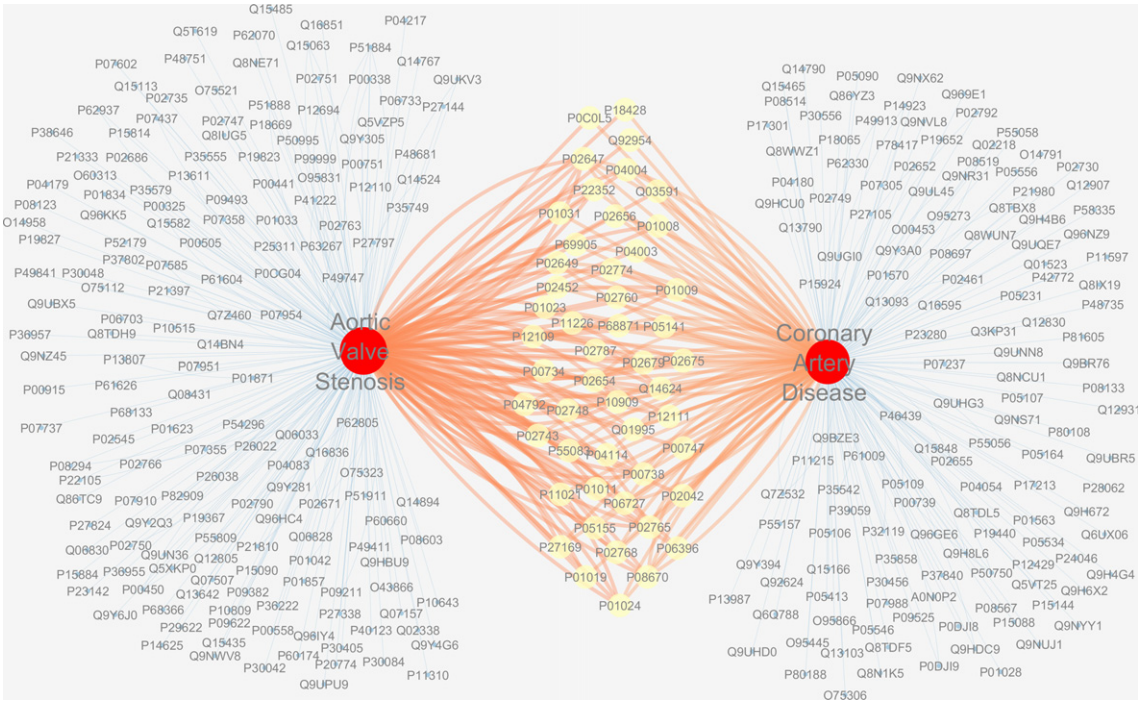

Fig. 2. Protein-disease associations network, after "Network Analysis". Node size and color were mapped to "Betweeness Centrality" (low values to small sizes and dark colors) and edge size and color were mapped to "Edge Betweeness" (low values to small sizes and dark colors). Nodes size and edges thickness are set according to the number of interactions for a given node. Big red central nodes refer to aortic valve stenosis and coronary artery disease. Larger yellow nodes refer to proteins linked to both conditions (non-specific markers). Smaller blue nodes represent proteins associated to only one of the conditions (surrogate markers). Proteins are represented with the respective UniProt code. The respective name can be found in Supplementary File 4.

non-conflicting up- or down-regulated proteins in either AVS or CAD. Alternatively, if the reader wants to perform step 6.1., raw data are available on Supplementary Files 1 and 2.

6.2. Run Cytoscape.

6.3. Go to "Apps" » "ClueGO" and run the plug-in. CluePedia will run automatically. Make sure you have previously installed both applications. To use ClueGO you will need to ask for a license key, which is free for academic use.

6.4. Set the "Analysis Mode" on the default "ClueGO: function" option.

6.5. Load the various protein lists in the gray boxes.

6.5.1. To add another box press the "+" sign under the box.

6.5.2. You may select each cluster color by clicking on the side color box.

6.6. Change the view style to "clusters" instead of "groups"; otherwise, you will see a network with a color pattern relative to groups of terms as defined by ClueGO using 'kappa' statistics (for further explanation, see the ClueGO manual at http://www.ici.upmc.fr/cluego/ClueGODocumentation.pdf), irrespective of the clusters determined a priori.

6.7. Define the ClueGO settings for the analysis:

6.7.1. Check the box relative to the biological process branch of GO (it is checked by default). Alternatively, you may check other GO branches or different pathways knowledgebases such as KEGG or Reactome.

6.7.2. Choose what kind of evidence you want to see. In this case, we selected "All", but one may see, for instance, only experimentally-derived annotations (for further information about evidence codes, see http://www.geneontology.org/page/guide-go-evidence-codes).

6.7.3. In a first analysis, do not change the network specificity. Although, depending on how complex the network shows up, one may drag the pointer towards left or right, in order to achieve a more global or detailed network, respectively. In such case, a new analysis should be performed. Hence, make sure to save the current analysis (step 6.14).

Note: The network specificity is defined according to the GO level. We used ClueGO's default settings: minimum GO level used was 3 and the maximum GO level used was 8. The settings can be changed to obtain a network of annotations with more global (GO level range: 1–4) or more specific terms (GO level range: 7–15). In the first case, ClueGO retrieves parent GO terms associated with a high number of genes, but with low level of granularity and generally lower relative frequency in the test gene set (because they annotate a higher proportion of the genome). On the opposite, in the second case, retrieved GO terms are associated with a low number of very specific genes, which result in higher frequencies in the test set, since they annotate only a minor, specific fraction of the genome.

It is also possible to precisely customize the desired GO levels in the "GO Tree Interval" field.

6.7.4. Choose the statistic test. We left the default option "two-sided hypergeometric test". Thus, the software will look at over and under-represented annotations. Still, it can be useful to run analysis with right-sided or left-sided tests to study over-represented mechanisms and repressed housekeeping processes in disease, respectively.

6.7.5. Define the Reference Set. We left the default "Selected Ontologies Reference Set", because we are comparing two conditions. Although, when using proteome data from samples of a specific organ/tissue/body fluid in a given pathophysiological condition, it would be preferable to use a "Custom Reference Set" (control samples) to avoid over-estimation of the significance of our annotated terms.

6.7.6. Check the box "Show only Pathways with pV ≤ 0.05", so that only significant biological processes are displayed.

6.8. Press the "start" button.

6.9. Adjust the style and view settings.

6.9.1. The network can be rearranged in the working space by simply mouse-selecting specific nodes and drag them to empty or less crowded spaces.
6.9.2. In order to align, scale and rotate the network or specific nodes go to “Layout” » “Scale” and a tool panel opens in the left side.
6.9.3. One can bypass the default color of nodes' labels to make them more readable. To achieve this, select the network or the desired nodes with the mouse and then, in the control panel, go to “Style” and right click on the square below “Byp.” column. Press “Set Bypass” and select the color of choice. Black color usually makes a good contrast with the label colors.
6.9.4. Go through the 3 different styles of node visualization (“View Style Settings”) to observe their specific features:
6.9.4.1. “Groups” style allows to visualize different groups of biological processes in different colors. Node size will be proportional to the process' significance.
6.9.4.2. “Significance” style shifts the interpretation of biological processes' significance from node size to a yellow-red-brown scale. Node size will be proportional to the number of mapped genes.
6.9.4.3. “Clusters” style (set in 6.5.) allows to visualize which biological processes belong to each cluster. Notice that processes more specifically associated to a defined cluster have the same color as the one you set on step 6.5.2. As in 6.9.4.1., node size will be proportional to the process' significance.

6.10. Analyze the network. You should check if the GO range defined in 6.7.3. is adequate to compare the clusters (e.g. sometimes lower GO levels mask potential biological differences between protein sets) and, if not, you should tune it to better visualize biological differences. Also, make sure to visualize the ClueGO Results tabs on the different clusters. You will find histograms whose bars' colors are defined according to the biological process and that depict the percentage of genes for each annotated term. Below histograms you have available pie-charts representing the most significant GO term per group and whose sections are correlated to the number of terms per group. Given that we set 4 clusters (up in AVS, down in AVS, up in CAD and down in CAD), we will find 4 corresponding tabs whose pie-charts describe specific up- and down-regulated phenomena in these conditions. Furthermore, an extra tab with unspecific terms is provided, meaning that they were not grouped in either of the four cases. Biologically, they represent undisturbed/house-keeping processes.
6.11. Export Network. Follow the same steps described in the sixth step of Cytoscape's tutorial (Fig. 3).
6.12. Export GO enrichment analysis results (tables and pie charts).
6.12.1. On the Table Panel, go to the “ClueGO Results” tab and press on the option “Save ClueGO Result Tables” (fifth icon) and “Save ClueGO Result Table as Excel Sheet” (sixth icon), in order to get the biological processes specifically associated with each disease, whether they are up- or down-regulated (Figs. 4 and 5).
6.12.2. If one finds pie-charts' subtitles difficult to read, one could, in each cluster tab, rotate the respective pie-chart by right clicking on it and selecting “Start” and then “Stop”. Such graphs give the user the distribution of biological processes potentially associated with a given cluster and, thus, they can be very informative with regard to the pathophysiological basis of the condition.

6.13. Perform CluePedia enrichment, if you want to add protein associations to the network, beyond the biological processes. This will add another layer of complexity to the network but can be very useful to understand graphically which proteins are related to the biological processes already represented, and to pinpoint which of those can be future therapeutic targets.
6.13.1. Go to “CluePedia” tab on the Table Panel.
6.13.2. Choose the option “Show genes that are from the initial clusters/added/enriched” (third icon). The software will incorporate more nodes representing the genes that were in the initial seed list and connect them to the discriminated biological processes.
6.13.3. Hand-select the non-gray nodes in ClueGO network output. These represent only those processes deregulated in AVS or CAD. If you have less gray nodes, you can handpick them and invert selection in “Select” menu – “Nodes” – “Invert Node Selection”, or simply pressing CTRL + I.
6.13.4. Go to “Select” » “Nodes” » “First Neighbors of Selected Nodes” » “Undirected”.
6.13.5. Press on the 13th icon under the Tools Menu to draw a “New Network From Selection (all edges)”. This will generate a new network that only contains deregulated biological processes as well as the genes/proteins associated to them.
Note: Pay attention to edge interpretation! Owing to CluePedia algorithm, which links genes/proteins to biological processes, there is the chance that some genes only associated to AVS or CAD are connected (edged) to deregulated processes in the other disease. This is because each gene can be associated to different GO terms. For instance, interleukin 6 (gene name: IL6, UniProt KB code: P05231), up-regulated in CAD, was connected to “monosaccharide biosynthetic process”, which is up-regulated in AVS. Though, this cytokine was not found elevated in AVS. This simply means that interleukin-6 also participates in such pathway.
6.13.6. Export the new network as in step 10
Note: It may require further adjustment of the view (step 6.9) (Fig. 6). Note that the network becomes more complex and difficult to read. This is because more nodes (genes encoding for the inputted proteins) and edges (association of the genes to the biological processes) were added.
6.13.7. Save this analysis as a new ClueGO session (next step)
Note: CluePedia functionalities go beyond protein-biological process annotation. It can also be useful to explore specific pathways by enriching the edges with different types of protein-protein interactions (e.g. activation, inhibition, binding) derived from experimental data or in silico data. Though, these features are beyond the scope of this tutorial and details can be found at http://www.ici.upmc.fr/cluepedia/CluePedia_Documentation.pdf.

6.14. Save session several times during the analysis and between analysis to avoid loss of previously performed analysis.
6.14.1. Go to “File” » “Save ClueGO Session as”
6.14.2. Name the file and choose its location.
6.14.3. Save the file as .cluego.

5. Integration of the outputs from bioinformatics analysis

After performing all the analysis described above, we could integrate the current knowledge on the diseases at scope, extract new biological questions and even point out some proteins that can be regarded as surrogate markers for these conditions. A detailed discussion of the

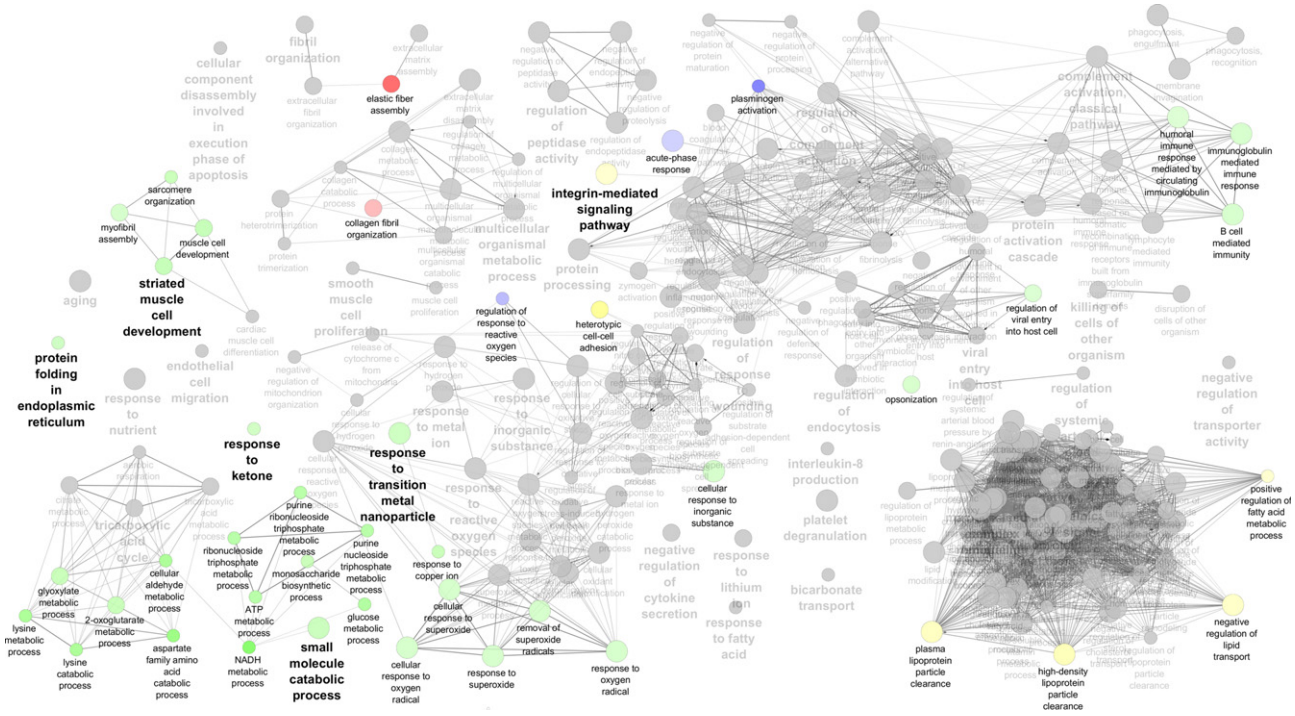

Fig. 3. ClueGO network depicting main biological processes associated to aortic valve stenosis and coronary artery disease. Gray-shaded nodes represent unspecific biological phenomena. Green-and red-shaded nodes represent, respectively, up- and down-regulated biological processes in aortic valve stenosis. Blue- and yellow-shaded nodes represent, respectively, up- and down-regulated processes in coronary artery disease. Node size is also proportional to the relevance of each biological phenomenon.

biological implications and hypotheses raised with this case-study can be found in Supplementary Materials. Herein, we will provide an overview of the main outputs from the bioinformatics tools so that the reader knows what to expect when performing similar analysis with different conditions. Furthermore, we will highlight the limitations of this approach and explain how to overcome them.

One of the first difficulties in performing this kind of bioinformatics analysis is to correctly assign a protein to its identifier. While UniProt KB accession code has becoming of generalized and consensual use, when one performs in-depth literature search and data-mining it is very probable that we come across with different codes that may even become obsolete (e.g. International Protein Index, IPI, that has been closed at September 2011 [22]). In those cases, we have to use UniProt KB to convert protein codes and resolve ambiguity of those unmapped through protein name (or features) correspondence. However, there will be always some proteins whose assignment will not be possible and, thus, they should not be considered for further analysis. Another limitation of the current approach is that we are gathering an enormous amount of data collected from different approaches (e.g. label-free liquid chromatography coupled to tandem mass spectrometry, quantitative mass spectrometry with tandem mass tags or imaging mass spectrometry), in different time-points and applying different statistical criterial either for protein identification (e.g. false discovery rate lower than 5% or 1%) or to test differential expression across groups (e.g. student *t*-test or Mann-Whitney *U* test). Furthermore, once databases are in constant actualization, it is expectable that the same study performed 5 or 10 years later would result in different outputs, that is, with different proteins identified. Therefore, one of the most important steps in performing this kind of data recycling is defining criteria for paper analysis and data-mining. As one can see in Supplementary Files 1 and 2, there is several criteria that one can use to filter our data and solve those issues. For instance, we can choose only studies performed with a similar approach by choosing the desired items in "Experimental Approach". Alternatively (or concomitantly), one may only choose proteins consistently deregulated across different methodologies or even only those that were validated (e.g. by western blot or ELISA approaches), applying a filter on "Validation". Also, during literature search, a filter to publication date can be used to get more homogenous

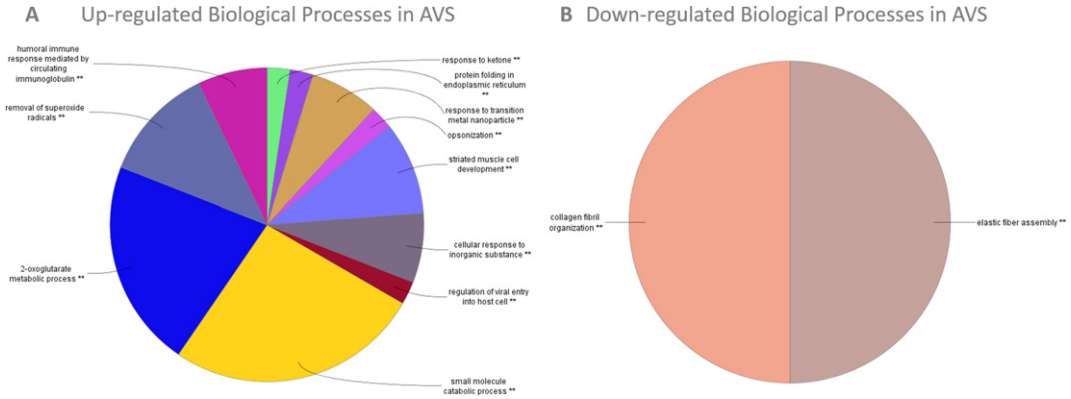

Fig. 4. Graphical representation of the specific and significant biological processes up-regulated (A) and down-regulated (B) in aortic valve stenosis.

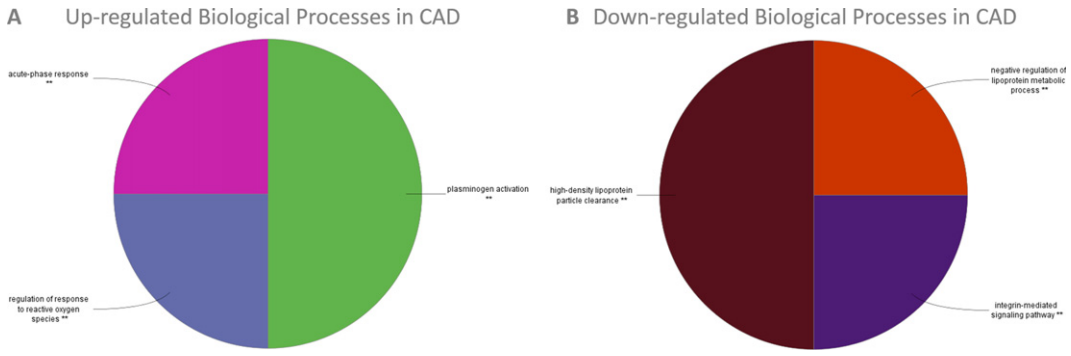

Fig. 5. Graphical representation of the specific and significant biological processes up-regulated (A) and down-regulated (B) in coronary artery disease.

datasets (e.g. studies from the last 5 years). To bypass the heterogeneity of statistical tests, one can select only proteins that have been found up- or down-regulated in more than one study. Thus, there is a large array of criteria that we can choose from to integrate data from different sources. However, data availability will have to weight on the initial decision-making step. Poorly known conditions will likely have a reduced number of papers published. In those cases, criteria must be less strict but, simultaneously, caution should be taken in establishing boundaries. For instance, samples collected from patients with co-morbidities that will change phenotype and, thus, the proteome, should be avoided. When data is scarce, the researcher should be the least stringent (that is, using all differentially expressed proteins across the different studies) and then validate new working hypothesis that may arise and candidate markers/therapeutic targets experimentally. In the present case-study, we decided to include all proteomics studies related to AVS and CAD, due to the relative low number of published articles available. Still, we have not considered those related to interventional or pharmacological therapies of any kind (this was our boundary). All studies enrolled used a false discovery rate no higher than 5% (or even 1%) as peptide/protein identification criteria. In the majority of the studies a student *t*-test was used to compare normal populations and a Mann-Whitney *U* test to compare non-normal populations. As the reader may see ahead, with this approach, some questions could be risen and some interesting proteins were found as surrogate markers for AVS and CAD (deserving future experimental validation). A detailed discussion of the biological insights and of the surrogate markers can be found in Supplementary Material. Herein, we will try to get an answer to the three questions formulated in Section 3.

After carefully selecting deregulated proteins in both conditions, we could get insight into the most representative biological processes deregulated in AVS and CAD, through STRING analysis (Table 4). Interestingly, AVS and CAD only shared "regulation of response to wounding" as one of the top 10 deregulated biological processes, probably reflecting the chronic nature of the defense response to stenotic and atherosclerotic lesions, respectively. Globally, we can see that in AVS there is a marked activation of proteases, complement and cellular transport, while in CAD hemostasis is disrupted and the activity of the immune system and lipoprotein metabolism are more pronounced. STRING's major limitation is the impossibility to separate a priori proteins in clusters in function of their differential tissue expression, biofluid level or disease association. Thus, rather than giving an integrated look over the up-regulated and down-regulated biological phenomena in both diseases, it essentially retrieves, at each time, which processes are implicated in each disease. Still, it should be highlighted that the latest version of STRING (10) is empowered with the association to DISEASES (http://diseases.jensenlab.org/) database making

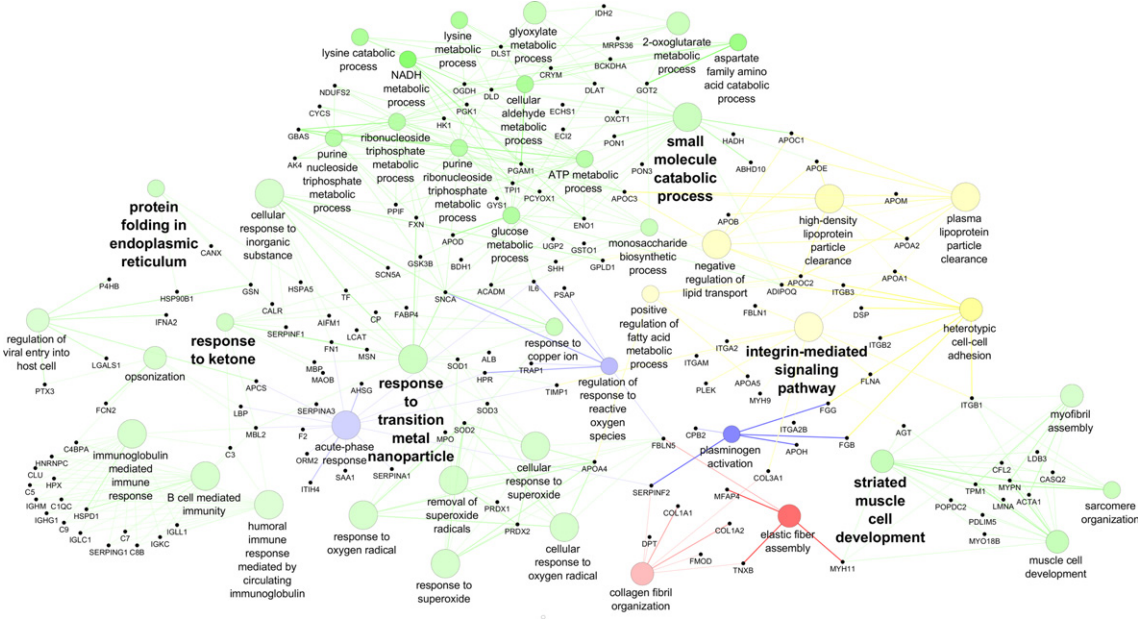

Fig. 6. ClueGO network after CluePedia-based enrichment. Besides the representation of the biological processes as in Fig. 3, the specific players (proteins) of such processes are highlighted with the respective gene name. Similarly, green and red nodes and edges represent, respectively, up- and down-regulation of biological processes or proteins in aortic valve stenosis. Blue and yellow nodes and edges represent, respectively, up- and down-regulation of biological processes or proteins in coronary artery disease. Some protein nodes are bi-colored due to double associations. Others are found uncolored as they have multiple associations and lack specificity.

possible to detect disease annotations [1]. Hence, it allows the observation of specific associations with other pathologies and, thus, the evaluation of their biomarker potential (see Top 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways in AVS and CAD in Table 4). For instance, AVS proteome dataset was associated to prion diseases, while CAD dataset was associated to pertussis.

With ClueGO, the user has the additional advantage of performing comparisons between different protein test sets, due to the cluster-based analysis (Fig. 3). ClueGO will colorize those nodes (at the protein or the term level) in function of their association to the condition and their regulation (up or down). This way, unspecific terms appear in gray shades and, the darker the node color, the stronger is the association of the protein/process to the condition. In Fig. 3, green and red nodes refer, respectively, to up- and down-regulated processes in AVS and, in turn, blue and yellow nodes refer to up- and down-regulated processes in CAD, respectively. By exporting the tables and pie charts (Figs. 4 and 5) of all clusters one can look, in detail, to the distribution of biological processes in each cluster and the *p*-value, indicating how strong the association of a given process to one cluster is. For example, we found a marked down-regulation of collagen fibril organization and elastic fiber assembly and a marked up-regulation of small molecule catabolism in AVS. In turn, we found that plasminogen activation is important in CAD pathogenesis, which is sided by a down-regulation of the signaling pathways mediated by integrin (Figs. 3 and 6). The biological implications of these findings can be found in Supplementary Material.

Hence, after going through STRING and ClueGO analyses, we can say that the answer to the first question formulated - is there any or some deregulated biological processes able to distinguish the pathogenesis or the phenotypical presentation of AVS and CAD? – is affirmative. However, while we get a general look over the biological phenomena with STRING, we can get specific annotations with ClueGO using higher GO range levels, helping us to improve the pathophysiological knowledge of these diseases. Unfortunately, the answer to the second question - do bone formation and metabolism-related biological processes have potential to distinguish both diseases? - is not totally positive. With regard to osteogenesis or bone formation, there was no direct GO term associated with any of the conditions, albeit calcification is a common event in AVS and CAD. One should recall, however, that such biological processes would likely become more evident if we only had selected and compared studies based on proteins extracted from valve or atheroma samples. Moreover, AVS proteome was associated with up-regulation of the response to transition metal nanoparticles and with the up-regulation of the cellular response to inorganic substance. Still, there was an association of small molecule catabolism (glycolytic metabolism) to AVS, deserving future scrutiny. It should be noted that STRING and other GO-based analysis software (such as ClueGO) rely on a set of known proteins/genes and medical terms. Thus, there is the chance that some phenomena, such as calcification, are not yet properly assigned to the conditions at scope. Nevertheless, the Gene Ontology Consortium is a project in constant actualization and it is the most complete information system available for protein function assignment by now, displaying >40,000 terms and over 400,000 annotations for *Homo sapiens* [23]. Ten years from now, rerunning these analyses in the same software would likely result in the observation of more deregulated processes, as the GO term library grows with new knowledge on protein function being reported and annotated every day.

Biomarker research is another field of major interest to the clinical practice, but it is also a never-ending task. Proteomics studies produce a high amount of data and it can be difficult to pinpoint which proteins can become potential markers. Fortunately, there are tools such as DisGeNET and Cytoscape that can help select and visualize the former. DisGeNET database sorts disease-associated proteins in descending order of score, meaning that those enlisted on top are more associated to the disease of matter. For that reason, Table 5 displays the Top 10 associated proteins for AVS and CAD, which are discussed in Supplementary Material. Notice that even the proteins with highest association to either of the diseases in scope are not specific to them. For instance, the DSI of enolase (the most specific protein in the Top10 of AVS-associated proteins) is 0.739 and the DSI of ribosome biogenesis protein WD repeat domain 12 (the most specific protein in the Top10 of CAD-associated proteins) is 0.854. Thus, both of these proteins fall on the wayside to meet the criteria for being a AVS or CAD biomarker. Furthermore, despite the existence of CAD-associated proteins with DSI = 1, these are poorly associated to the pathology, with scores below 0.005, thus they should not be considered. Therefore, to answer the last

Table 4
Top 10 biological processes and Top 10 deregulated pathways in Aortic Valve Stenosis and Coronary Artery Disease and respective false discovery rate (FDR), indicated by STRING analysis.

Aortic valve stenosis		Coronary artery disease	
GO biological process	FDR	GO biological process	FDR
Protein activation cascade	2.14×10^{-12}	Wound healing	6.58×10^{-20}
Regulation of proteolysis	1.17×10^{-10}	Blood coagulation	3.59×10^{-18}
Complement activation	1.46×10^{-10}	Response to wounding	3.59×10^{-18}
Single-organism metabolic process	1.46×10^{-10}	Plasma lipoprotein particle remodeling	3.59×10^{-18}
Regulation of peptidase activity	1.62×10^{-10}	Response to stress	2.08×10^{-17}
Single-organism catabolic process	2.9×10^{-10}	Regulation of body fluid levels	3.7×10^{-17}
Regulation of response to wounding	2.9×10^{-10}	Regulation of biological quality	8.91×10^{-17}
Vesicle-mediated transport	3.18×10^{-10}	Regulation of response to wounding	1.19×10^{-16}
Endocytosis	1.33×10^{-9}	Regulation of plasma lipoprotein particle levels	5.25×10^{-16}
Receptor-mediated endocytosis	4.71×10^{-9}	Regulation of immune system process	1.59×10^{-15}

Aortic valve stenosis		Coronary artery disease	
KEGG pathways	FDR	KEGG pathways	FDR
Complement and coagulation cascades	2.94×10^{-12}	Complement and coagulation cascades	6.64×10^{-17}
Carbon metabolism	1.11×10^{-7}	Pertussis	1.44×10^{-5}
Prion diseases	4.26×10^{-5}	*Staphylococcus aureus* infection	1.89×10^{-5}
Glycolysis/Gluconeogenesis	4.27×10^{-5}	PPAR signaling pathway	1.06×10^{-4}
Fatty acid degradation	1.41×10^{-3}	Glutathione metabolism	1.74×10^{-4}
Valine, leucine and isoleucine degradation	1.41×10^{-3}	Legionellosis	2.32×10^{-4}
Butanoate metabolism	2.19×10^{-3}	ECM-receptor interaction	2.56×10^{-4}
Staphylococcus aureus infection	2.19×10^{-3}	African trypanosomiasis	2.56×10^{-4}
Citrate cycle (TCA cycle)	2.46×10^{-3}	Platelet activation	2.9×10^{-4}
Glycine, serine and threonine metabolism	6.37×10^{-3}	Renin-angiotensin system	2.9×10^{-4}

Abbreviations: ECM: extracellular matrix; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPAR: peroxisome proliferator-activated receptor; TCA: tricarboxylic acids.

question - is there any protein or group of proteins that may become surrogate markers for anticipated diagnosis, prognosis or therapeutic monitoring of AVS or CAD? - we performed a Cytoscape analysis, followed by a DisGeNET search. As one can see in Figs. 1 and 2, with Cytoscape we can detain in a single analysis and network which proteins are potentially associated with AVS and CAD (small blue nodes in Fig. 2) and those who do not exhibit specific associations (large yellow nodes in Fig. 2). From this analysis, 50 proteins (yellow nodes) were found to be simultaneously linked to AVS and CAD, while the remaining 299 proteins (blue nodes) were found exclusively associated with AVS (164) or CAD (135). Interestingly, none of the unique proteins indicated by Cytoscape were among DisGeNET's Top 10 of disease-associated proteins list (Table 5, Fig. 7A and B). Despite the fact that 142 (out of 164) proteins had not been associated to AVS and 106 (out of 135) proteins had not been associated to CAD (Fig. 7C), according to Cytoscape and DisGeNET, the majority of them were already linked with certain diseases. Still, there are proteins with biomarker potential. For instance, the popeye domain containing protein 2 (Q9HBU9), a protein essentially expressed in heart muscle, was not found associated to any disease in DisGeNET and was found up-regulated in right ventricle biopsies collected from AVS patients [6]. In the same way, the 28S ribosomal protein S36 (mitochondrial, P82909), has not yet been assigned to any particular condition, but AVS, as reported in the previous study [6]. To name one last example, the E3 ubiquitin-protein ligase complex Ankyrin repeat and SOCS box protein 7 (Q9H672) is yet to be associated to a pathological setting and was found exclusively up-regulated in sera of CAD patients [24]. These and other examples of potential markers (octagonal orange nodes) for AVS or CAD are summarized in Fig. 8 (a new network built from the information retrieved from ClueGO and DisGeNET). Except for the 28S ribosomal protein S36 and the popeye domain containing protein 2, the remaining potential markers were not significantly associated with biological processes up- or down-regulated in AVS or CAD. This may be a consequence of the GO range used in ClueGO analysis or the lower percentage of genes involved in biological phenomena that were present in each cluster. Though, they should not be disregarded because none of those were found to have disease associations described to date according to DisGeNET. Furthermore, even acknowledging that the remaining proteins in Fig. 8 (identified through the coding gene) could be associated to other conditions, there is still a large set of proteins whose biomarker value – either alone or in multiplex – remain to validate in studies with large cohorts. Thus, the answer for this last question is also affirmative.

6. Final remarks

Herein we describe the major steps to perform bioinformatics analysis of proteomic data collected from available studies in the literature. We took advantage of a clinical problem – to distinguish the pathological features of AVS and CAD and the need to pinpoint surrogate markers – to take a tour over 4 very useful tools that can help guide research through new working hypothesis, promoting data recycling and integration of knowledge produced worldwide. We started to compare annotated biological processes according to specific deregulated proteins in AVS and CAD and we saw that STRING performs a rapid GO enrichment analysis with a list of proteins from one of the conditions. Thus, it is recommended to execute a STRING analysis whenever the user wants to get the idea of the main biological phenomena implicated in a given pathological setting. STRING can also be relevant in finding indirect markers not present in the initial dataset, by adding second shell interactors to the network. Although, if one looks to compare specifically two or more conditions or a condition with a healthy status, ClueGO would be more helpful. Even though ClueGO analysis is time-consuming and requires large computer memory, final results provide specific associations of biological processes (or molecular functions, cellular components or pathways) to each one of the clusters defined a priori. These associations may need to be validated experimentally (if not available in literature), but they can help to design novel hypothesis by identifying specific markers. Antibody-based tests or mass spectrometry-based targeted proteomics are commonly used experimental approaches that can help such validation.

Then, we looked over disease molecular fingerprints with the help of Cytoscape and DisGeNET, a large database comprising disease-gene (or disease-protein) associations. As we could see, DisGeNET can be useful whenever the goal is to check if a certain deregulated protein in a given pathophysiological context was previously associated to a disease, helping with the selection of truly specific disease markers. With this tool, we have seen that aside from angiotensin-converting enzyme, there are 9 other different associated proteins with AVS and 9 more for CAD in DisGeNET's Top 10 list. Nonetheless, those proteins cannot be used as single markers for AVS or CAD owing to their connection to other diseases. Therefore, they could be either tested in multiplex for diagnosis/prognosis of AVS or CAD or we can resort to Cytoscape's single-associated proteins (smaller blue nodes in Fig. 2 and octagonal orange nodes in Fig. 8) to check for potential markers of AVS or CAD, that have not known associations to other conditions. Indeed, from the set of AVS-linked proteins we could find surrogate markers for AVS (e.g. popeye domain containing protein 2 and 28S ribosomal protein S36, mitochondrial). Likewise, from the set of CAD-linked protein we were able to detect potential markers for CAD, for instance, the Ankyrin repeat and SOCS box protein 7. Such proteins are not yet associated to any disease, according to DisGeNET, and, thus, their marker potential deserves to be tested.

Thereby, with this paper we have outlined an example of use of publicly available bioinformatics resources for data integration of

Table 5
Top 10 associated proteins with Aortic Valve Stenosis and Coronary Artery Disease, respective DisGeNET score and Disease-Specificity Index (DSI).

Aortic valve stenosis				Coronary artery disease			
DisGeNET code: C0003507				DisGeNET code: C1956346			
UniProt code	Protein name	Score	DSI	UniProt code	Protein name	Score	DSI
P23582	C-type natriuretic peptide	0.08300	0.594	P12821	Angiotensin-converting enzyme	0.26526	0.334
P13929	Beta-enolase	0.08000	0.739	Q9C0D0	Phosphatase and actin regulator 1	0.25265	0.723
P12821	Angiotensin-converting enzyme	0.01118	0.334	P16442	Histo-blood group ABO system transferase	0.24846	0.466
P16860	Natriuretic peptides B	0.00572	0.554	O14807	Ras-related protein M-Ras	0.24819	0.796
P02649	Apolipoprotein E	0.00563	0.332	O14495	Phospholipid phosphatase 3	0.24792	0.769
P50052	Type-2 angiotensin II receptor	0.00536	0.515	Q9UKP4	A disintegrin and metalloproteinase with thrombospondin motifs 7	0.24636	0.731
Q92833	Protein Jumonji	0.00300	0.691	Q9GZL7	Ribosome biogenesis protein WD Repeat Domain 12	0.24473	0.854
P04114	Apolipoprotein B-100	0.00300	0.462	Q86WB0	Nuclear-interacting partner of anaplastic lymphoma receptor tyrosine kinase	0.24318	0.796
P22301	Interleukin-10	0.00291	0.296	Q9UQQ2	SH2B adapter protein 3	0.24291	0.608
Q8WXI7	Mucin-16	0.00272	0.562	O43680	Transcription factor 21	0.24237	0.680

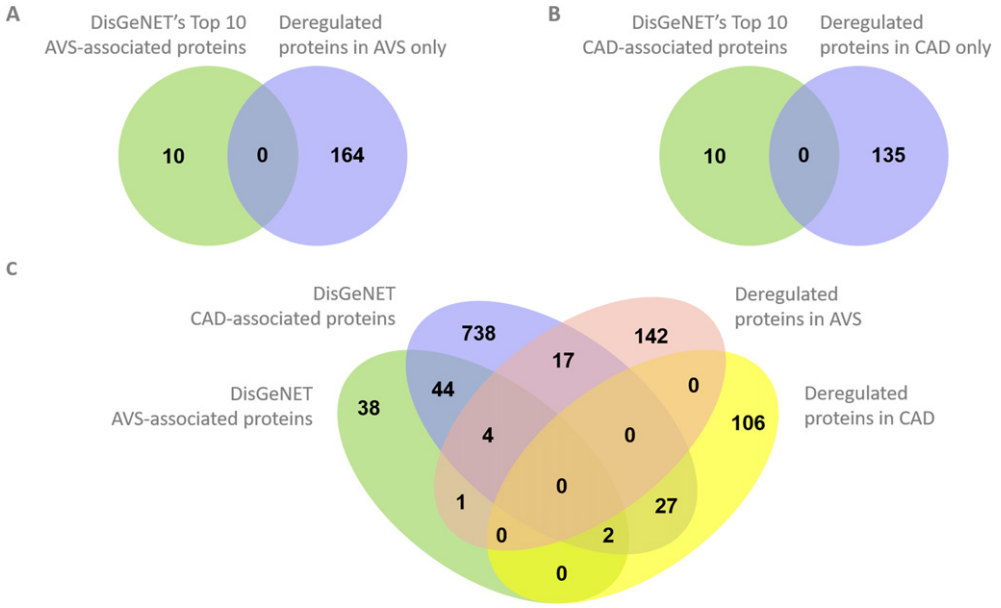

Fig. 7. Venn diagrams showing the absence of specific aortic valve stenosis (AVS) proteins in DisGeNET's Top 10 list for AVS (A) as well as the absence of specific coronary artery disease (CAD) proteins in DisGeNET's Top 10 list for CAD (B). Even crossing disease-specific proteins with the complete DisGeNET's protein list for each conditions, there is a substantial number of proteins (142 for AVS and 106 for CAD) yet to validate as AVS- or CAD-associated (C).

proteomics studies, producing a snapshot of the current knowledge in a particular disease setting. These results can be used to focus research on loose ends and to avoid work duplication.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2017.03.015.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through UnIC, iBiMED, QOPNA research units (UID/IC/00051/2013, UID/BIM/04501/2013, PEst-C/QUI/UI0062/2013), by project DOCnet (NORTE-01-0145-FEDER-000003), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). Rui Vitorino and Fábio Trindade are supported by individual grants (IF/00286/2015 and SFRH/BD/111633/

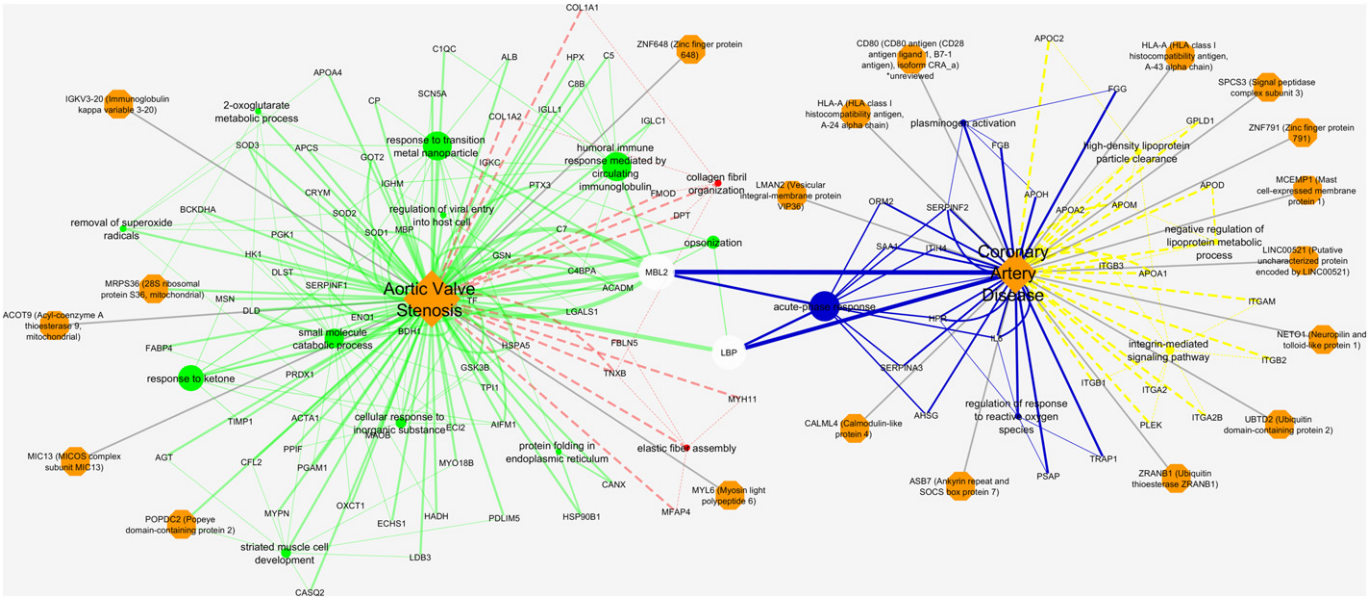

Fig. 8. Cytoscape network summarizing the results of the bioinformatics analysis. Aortic valve stenosis (AVS) and coronary artery disease (CAD) appear as two central nodes (orange diamonds) and specific associations to biological processes and proteins (represented by the coding genes) are represented by edges. Green solid edges and red dashed edges represent, respectively, up- and down-regulation in AVS. Blue solid edges and yellow dashed edges represent, respectively, up- and down-regulation in CAD. Octagonal orange nodes represent specific proteins found to be exclusively associated to AVS or CAD but without any known association to other pathologies, according to DisGeNET.

2015, respectively). The authors would like to thank the generous and precious help of Dr. Pablo Porras, from the European Bioinformatics Institute (EMBL-EBI), in the technical revision of the tutorial.

References

[1] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) D447–D452, http://dx.doi.org/10.1093/nar/gku1003.

[2] J. Piñero, N. Queralt-Rosinach, À. Bravo, J. Deu-Pons, A. Bauer-Mehren, M. Baron, F. Sanz, L.I. Furlong, DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database (2015)http://dx.doi.org/10.1093/database/bav028.

[3] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, http://dx.doi.org/10.1101/gr.1239303.

[4] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25 (2009) 1091–1093, http://dx.doi.org/10.1093/bioinformatics/btp101.

[5] C. von Mering, L.J. Jensen, B. Snel, S.D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M.A. Huynen, P. Bork, STRING: known and predicted protein–protein associations, integrated and transferred across organisms, Nucleic Acids Res. 33 (2005) D433–D437, http://dx.doi.org/10.1093/nar/gki005.

[6] B. Littlejohns, K. Heesom, G.D. Angelini, M.-S. Suleiman, The effect of disease on human cardiac protein expression profiles in paired samples from right and left ventricles, Clin. Proteomics 11 (2014) 1–14, http://dx.doi.org/10.1186/1559-0275-11-34.

[7] M.R. Dweck, N.A. Boon, D.E. Newby, Calcific aortic stenosis: a disease of the valve and the myocardium, J. Am. Coll. Cardiol. 60 (2012) 1854–1863, http://dx.doi.org/10.1016/j.jacc.2012.02.093.

[8] B.R. Hughes, G. Chahoud, J.L. Mehta, Aortic stenosis: is it simply a degenerative process or an active atherosclerotic process? Clin. Cardiol. 28 (2005) 111–114, http://dx.doi.org/10.1002/clc.4960280303.

[9] I. Falcão-Pires, C. Gavina, A.F. Leite-Moreira, Understanding the molecular and cellular changes behind aortic valve stenosis, Curr. Pharm. Biotechnol. 13 (2012) 2485–2496, http://dx.doi.org/10.2174/138920112804583050.

[10] J.A. Finegold, P. Asaria, D.P. Francis, Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations, Int. J. Cardiol. 168 (2013) 934–945, http://dx.doi.org/10.1016/j.ijcard.2012.10.046.

[11] F. Otsuka, S. Yasuda, T. Noguchi, H. Ishibashi-Ueda, Pathology of coronary atherosclerosis and thrombosis, Cardiovasc. Diagn. Ther. 6 (2016) 396–408, http://dx.doi.org/10.21037/cdt.2016.06.01.

[12] E. Yetkin, J. Waltenberger, Molecular and cellular mechanisms of aortic stenosis, Int. J. Cardiol. 135 (2009) 4–13, http://dx.doi.org/10.1016/j.ijcard.2009.03.108.

[13] G. Christodoulidis, T.J. Vittorio, M. Fudim, S. Lerakis, C.E. Kosmas, Inflammation in coronary artery disease, Cardiol. Rev. 22 (2014) 279–288, http://dx.doi.org/10.1097/CRD.0000000000000006.

[14] G.K. Hansson, Inflammation, atherosclerosis, and coronary artery disease, N. Engl. J. Med. 352 (2005) 1685–1695, http://dx.doi.org/10.1056/NEJMra043430.

[15] P. Libby, P. Theroux, Pathophysiology of coronary artery disease, Circulation 111 (2005) 3481–3488http://circ.ahajournals.org/content/111/25/3481.abstract.

[16] T. Martín-Rojas, F. Gil-Dones, L.F. Lopez-Almodovar, L.R. Padial, F. Vivanco, M.G. Barderas, Proteomic profile of human aortic stenosis: insights into the degenerative process, J. Proteome Res. 11 (2012) 1537–1550, http://dx.doi.org/10.1021/pr2005692.

[17] I. Ikonomidis, C.A. Michalakeas, J. Parissis, I. Paraskevaidis, K. Ntai, I. Papadakis, M. Anastasiou-Nana, J. Lekakis, Inflammatory markers in coronary artery disease, Biofactors 38 (2012) 320–328, http://dx.doi.org/10.1002/biof.1024.

[18] D.A. Lerman, S. Prasad, N. Alotti, Calcific aortic valve disease: molecular mechanisms and therapeutic approaches, Eur. Cardiol. 10 (2015) 108–112, http://dx.doi.org/10.15420/ecr.2015.10.2.108.

[19] A. Poduri, A. Bahl, K.K. Talwar, M. Khullar, Proteomic analysis of circulating human monocytes in coronary artery disease, Mol. Cell. Biochem. 360 (2012) 181–188, http://dx.doi.org/10.1007/s11010-011-1055-3.

[20] M.-Y. Lee, C.-H. Huang, C.-J. Kuo, C.-L.S. Lin, W.-T. Lai, S.-H. Chiou, Clinical proteomics identifies urinary CD14 as a potential biomarker for diagnosis of stable coronary artery disease, PLoS One 10 (2015), e0117169. http://dx.doi.org/10.1371/journal.pone.0117169.

[21] T. Martin-Rojas, L. Mourino-Alvarez, S. Alonso-Orgaz, E. Rosello-Lleti, E. Calvo, L.F. Lopez-Almodovar, M. Rivera, L.R. Padial, J.A. Lopez, F. de la Cuesta, M.G. Barderas, iTRAQ proteomic analysis of extracellular matrix remodeling in aortic valve disease, Sci. Rep. 5 (2015) 17290http://dx.doi.org/10.1038/srep17290.

[22] J. Griss, M. Martín, C. O'Donovan, R. Apweiler, H. Hermjakob, J.A. Vizcaíno, Consequences of the discontinuation of the international protein index (IPI) database and its substitution by the UniProtKB "complete proteome" sets, Proteomics 11 (2011) 4434–4438, http://dx.doi.org/10.1002/pmic.201100363.

[23] Several-Authors, Gene Ontology Consortium: going forward, Nucleic Acids Res. 43 (2015) D1049–D1056, http://dx.doi.org/10.1093/nar/gku1179.

[24] Y. Han, S. Zhao, Y. Gong, G. Hou, X. Li, L. Li, Serum cyclin-dependent kinase 9 is a potential biomarker of atherosclerotic inflammation, Oncotarget 7 (2) (2015)http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&amp.

[25] F. Gil-Dones, T. Martín-Rojas, L.F. López-Almodovar, R. Juárez-Tosina, F. de La Cuesta, G. Álvarez-Llamas, S. Alonso-Orgaz, F. Vivanco, L. Rodríguez-Padial, M.G. Barderas, Development of an optimal protocol for the proteomic analysis of stenotic and healthy aortic valves, Rev. Española Cardiol. 63 (2010) 46–53 (English Ed).

[26] G. Alvarez-Llamas, T. Martín-Rojas, F. de la Cuesta, E. Calvo, F. Gil-Dones, V.M. Dardé, L.F. Lopez-Almodovar, L.R. Padial, J.-A. Lopez, F. Vivanco, M.G. Barderas, Modification of the secretion pattern of proteases, inflammatory mediators, and extracellular matrix proteins by human aortic valve is key in Severe aortic stenosis, Mol. Cell. Proteomics 12 (2013) 2426–2439, http://dx.doi.org/10.1074/mcp.M113.027425.

[27] K. Satoh, K. Yamada, T. Maniwa, T. Oda, K. Matsumoto, Monitoring of serial Presurgical and postsurgical changes in the serum proteome in a series of patients with calcific aortic stenosis, Dis. Markers 2015 (2015).

[28] F. Gil-Dones, T. Martin-Rojas, L.F. Lopez-Almodovar, F. de la Cuesta, V.M. Darde, G. Alvarez-Llamas, R. Juarez-Tosina, G. Barroso, F. Vivanco, L.R. Padial, M.G. Barderas, Valvular aortic stenosis: a proteomic insight, Clin. Med. Insights Cardiol. 4 (2010) 1–7, http://dx.doi.org/10.4137/CMC.S3884.

[29] F. Gil-Dones, V.M. Darde, S. Alonso-Orgaz, L.F. Lopez-Almodovar, L. Mourino-Alvarez, L.R. Padial, F. Vivanco, M.G. Barderas, Inside human aortic stenosis: a proteomic analysis of plasma, J. Proteomics 75 (2012) 1639–1653, http://dx.doi.org/10.1016/j.jprot.2011.11.036.

[30] H. Suzuki, M. Chikada, M.K. Yokoyama, M.S. Kurokawa, T. Ando, H. Furukawa, M. Arito, T. Miyairi, T. Kato, Aberrant glycosylation of Lumican in aortic valve stenosis revealed by proteomic analysis, Int. Heart J. 57 (2016) 104–111, http://dx.doi.org/10.1536/ihj.15-252.

[31] L. Mourino-Alvarez, I. Iloro, F. de la Cuesta, M. Azkargorta, T. Sastre-Oliva, I. Escobes, L.F. Lopez-Almodovar, P.L. Sanchez, H. Urreta, F. Fernandez-Aviles, A. Pinto, L.R. Padial, F. Akerström, F. Elortza, M.G. Barderas, MALDI-imaging mass spectrometry: a step forward in the anatomopathological characterization of stenotic aortic valve tissue, Sci. Rep. 6 (2016) 27106http://dx.doi.org/10.1038/srep27106.

[32] K.-I. Matsumoto, K. Satoh, T. Maniwa, A. Araki, R. Maruyama, T. Oda, Noticeable decreased expression of tenascin-X in calcific aortic valves, Connect. Tissue Res. 53 (2012) 460–468.

[33] S. Krishnan, J. Huang, H. Lee, A. Guerrero, L. Berglund, E. Anuurad, C.B. Lebrilla, A.M. Zivkovic, Combined high-density lipoprotein proteomic and Glycomic profiles in patients at risk for coronary artery disease, J. Proteome Res. 14 (2015) 5109–5118, http://dx.doi.org/10.1021/acs.jproteome.5b00730.

[34] T. Basak, V.S. Tanwar, G. Bhardwaj, N. Bhardwaj, S. Ahmad, G. Garg, S.V.G. Karthikeyan, S. Seth, S. Sengupta, Plasma proteomic analysis of stable coronary artery disease indicates impairment of reverse cholesterol pathway, Sci. Rep. 6 (2016), 28042. http://dx.doi.org/10.1038/srep28042.

[35] L. Yan, D. Wang, H. Liu, X. Zhang, H. Zhao, L. Hua, P. Xu, Y. Li, A pro-Atherogenic HDL profile in coronary heart disease patients: an iTRAQ labelling-based proteomic approach, PLoS One 9 (2014), e98368. http://dx.doi.org/10.1371%252Fjournal.pone.0098368.

[36] M. Riwanto, L. Rohrer, B. Roschitzki, C. Besler, P. Mocharla, M. Mueller, D. Perisa, K. Heinrich, L. Altwegg, A. von Eckardstein, T.F. Lüscher, U. Landmesser, Altered activation of endothelial anti- and Proapoptotic pathways by high-density lipoprotein from patients with coronary artery DiseaseClinical perspective, Circulation 127 (2013) 891–904http://circ.ahajournals.org/content/127/8/891.abstract.

[37] A. Salgado-Somoza, E. Teijeira-Fernández, Á.L. Fernández, J.R. González-Juanatey, S. Eiras, Changes in lipid transport-involved proteins of epicardial adipose tissue associated with coronary artery disease, Atherosclerosis 224 (2012) 492–499, http://dx.doi.org/10.1016/j.atherosclerosis.2012.07.014.

[38] M.P. Donahue, K. Rose, D. Hochstrasser, J. Vonderscher, P. Grass, S.-D. Chibout, C.L. Nelson, P. Sinnaeve, P.J. Goldschmidt-Clermont, C.B. Granger, Discovery of proteins related to coronary artery disease using industrial-scale proteomics analysis of pooled plasma, Am. Heart J. 152 (2006) 478–485, http://dx.doi.org/10.1016/j.ahj.2006.03.007.

[39] R.K. Vangala, V. Ravindran, K. Kamath, V.S. Rao, H. Sridhara, Novel network biomarkers profile based coronary artery disease risk stratification in Asian Indians, Adv. Biomed. Res. 2 (2013).

[40] L.U. Zimmerli, E. Schiffer, P. Zürbig, D.M. Good, M. Kellmann, L. Mouls, A.R. Pitt, J.J. Coon, R.E. Schmieder, K.H. Peter, H. Mischak, W. Kolch, C. Delles, A.F. Dominiczak, Urinary proteomic biomarkers in coronary artery disease, Mol. Cell. Proteomics 7 (2008) 290–298, http://dx.doi.org/10.1074/mcp.M700394-MCP200.

[41] O.B. Bleijerveld, P. Wijten, S. Cappadona, E.A. McClellan, A.N. Polat, R. Raijmakers, J.-W. Sels, L. Colle, S. Grasso, H.W. van den Toorn, B. van Breukelen, A. Stubbs, G. Pasterkamp, A.J.R. Heck, I.E. Hoefer, A. Scholten, Deep proteome profiling of circulating granulocytes reveals bactericidal/permeability-increasing protein as a biomarker for Severe atherosclerotic coronary stenosis, J. Proteome Res. 11 (2012) 5235–5244, http://dx.doi.org/10.1021/pr3004375.

[42] C. von zur Muhlen, E. Schiffer, P. Zuerbig, M. Kellmann, M. Brasse, N. Meert, R.C. Vanholder, A.F. Dominiczak, Y.C. Chen, H. Mischak, C. Bode, K. Peter, Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients, J. Proteome Res. 8 (2009) 335–345, http://dx.doi.org/10.1021/pr800615t.

[43] C. Banfi, M. Brioschi, G. Marenzi, M. De Metrio, M. Camera, L. Mussoni, E. Tremoli, Proteome of platelets in patients with coronary artery disease, Exp. Hematol. 38 (2010) 341–350, http://dx.doi.org/10.1016/j.exphem.2010.03.001.

[44] T. Vaisar, S. Pennathur, P.S. Green, S.A. Gharib, A.N. Hoofnagle, M.C. Cheung, J. Byun, S. Vuletic, S. Kassim, P. Singh, H. Chea, R.H. Knopp, J. Brunzell, R. Geary, A. Chait, X.-Q. Zhao, K. Elkon, S. Marcovina, P. Ridker, J.F. Oram, J.W. Heinecke, Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL, J. Clin. Invest. 117 (2007) 746–756, http://dx.doi.org/10.1172/JCI26206.

[45] S.-A. You, S.R. Archacki, G. Angheloiu, C.S. Moravec, S. Rao, M. Kinter, E.J. Topol, Q. Wang, Proteomic approach to coronary atherosclerosis shows ferritin light chain as a significant marker: evidence consistent with iron hypothesis in atherosclerosis, Physiol. Genomics 13 (2003) 25–30http://physiolgenomics.physiology.org/content/13/1/25.abstract.

[46] T. Vaisar, P. Mayer, E. Nilsson, X.-Q. Zhao, R. Knopp, B.J. Prazen, HDL in humans with cardiovascular disease exhibits a proteomic signature, Clin. Chim. Acta. 411 (2010) 972–979, http://dx.doi.org/10.1016/j.cca.2010.03.023.

[47] F. de la Cuesta, G. Alvarez-Llamas, A.S. Maroto, A. Donado, I. Zubiri, M. Posada, L.R. Padial, A.G. Pinto, M.G. Barderas, F. Vivanco, A proteomic focus on the alterations occurring at the human atherosclerotic coronary intima, Mol. Cell. Proteomics 10 (2011)http://dx.doi.org/10.1074/mcp.M110.003517.

[48] S. Cooksley-Decasper, H. Reiser, D.S. Thommen, B. Biedermann, M. Neidhart, J. Gawinecka, G. Cathomas, F.C. Franzeck, C. Wyss, R. Klingenberg, P. Nanni, B. Roschitzki, C. Matter, P. Wolint, M.Y. Emmert, M. Husmann, B. Amann-Vesti, W. Maier, S. Gay, T.F. Lüscher, A. von Eckardstein, D. Hof, Antibody phage display assisted identification of junction Plakoglobin as a potential biomarker for atherosclerosis, PLoS One 7 (2012), e47985. http://dx.doi.org/10.1371%252Fjournal.pone.0047985.

[49] M. Ghatge, A. Sharma, R.K. Vangala, Association of γ-glutamyl transferase with premature coronary artery disease, Biomed. Rep. 4 (2016) 307–312.