Austin Journal of Proteomics, Bioinformatics & Genomics

Open Access

Austin Publishing Group

Research Article

Network Biology Analysis of the Human Abdominal Aortic Aneurysm Plasma Proteome

Alessandra Modesti[1]*, Tiziana Alberio[2], Tania Gamberi[1], Mauro Fasano[2], Francesca Magherini[1], Tania Fiaschi[1], Manuela Balzi[1], Merry L. Lindsey[3] and Pietro A. Modesti[4]*

[1]Dipartimento di Scienze Biomediche, Sperimentali e Cliniche, Università degli Studi di, Firenze, Italy

[2]Dipartimento di Scienze Teoriche e Applicate, Biomedical Research Division, Università, dell'Insubria, Italy

[3]Mississippi Center for Heart Research, Department of Physiology and Biophysics, University of Mississippi Medical Center, and Research Service, G.V. (Sonny) Montgomery,Veterans Affairs Medical Center, Jackson, MS

[4]Dipartimento di Medicina Sperimentale e Clinica, Università degli Studi di Firenze, Italy

***Corresponding author:** Alessandra Modesti - Dipartimento di Scienze Biomediche, Sperimentali e Cliniche,Viale G. Morgagni 50–50134. Università degli Studi di Firenze, Italy, Fax: +390554598905; E-mail: alessandra.modesti@unifi.it

Pietro Amedeo Modesti – Dipartimento di Medicina Sperimentale e Clinica Largo Brambilla, 3-50134, Università degli Studi di Firenze, Italy. Fax: +39 0554378638; E-mail: pa.modesti@unifi.it

Received: September 30, 2014; **Accepted:** November 24, 2014; **Published:** November 26, 2014

Abstract

Objective: Abdominal aortic aneurysm (AAA) affects up to 9% of adults, and the incidence of asymptomatic and ruptured AAA currently accounts for 1–2% of male deaths. A number of studies have applied classical proteomics methodology to identify plasma biomarkers of AAA.

The use of a systems biology protein-protein interaction network analysis on proteomics results can reveal novel mechanisms.

Methods: We performed a bioinformatics analysis on a compiled set of proteins previously identified in multiple proteomics studies. Results from a total of nine papers were identified and included in the analysis. The 64 proteins related to AAA were first analyzed by over-representation analysis (ORA) using Webgestalt. Afterwards, the same input list was submitted to BioProfiling. This portal used statistical methodology for the network based interpretation of the protein list (Rspider, PPIspider). The networks obtained by both Reactome and PPI analysis were further analyzed by ORA using the JEPETTO application in the Cytoscape environment.

Results: This analysis revealed a strong over-representation for proteins related not only to blood clotting but also to cellular mediated response, such as cell adhesion and cytokine activation of platelets and white blood cells.

Conclusions: We used a network strategy to generate statistically valid novel hypotheses about biological mechanisms related to AAA, which may be useful in providing new insights into the understanding of the pathogenesis of AAA.

Keywords: Abdominal aortic aneurysm; Proteomics; Network analysis

Introduction

Abdominal aortic aneurysm (AAA) affects up to 9% of adults older than 65 years of age [1]. The mortality rate for patients with ruptured AAA approaches 90%, with 40–50% of deaths occurring before the patient reaches the hospital. Asymptomatic AAA is easily detectable with abdominal ultrasound, and screening programs are now considered a reasonable practice option since surgical repair is restricted to AAA >5.5 cm at centers with acceptable operative mortality [2-6]. Identifying markers other than AAA size that would predict a risk of rupture is the focus of current investigation. For AAA, one significant limitation relates to the fact that many biomarkers are not disease specific; indeed most of them are also markers for atherosclerosis [7]. Serum Elastin Peptide holds promise to be an improved biomarker to predict AAA expansion and rupture [8-10] because it positively correlates between its plasma level and the AAA growth rate [10]. The extent of aortic dilation and vessel tortuosity is associated with blood levels of several fibrinolytic factors, including D-dimer, fibrinogen/fibrin degradation products, and plasmin inhibitor-plasmin complexes [11]. Specifically, there is a positive correlation between AAA size and fibrinogen concentration in the plasma [12]. For AAA many of the identified differences in the serum proteome often represented proteins highly abundant in serum rather than true candidate biomarkers [13]. Using a proteomic approach, our team has confirmed several plasma biomarkers that others have previously identified as being associated with AAA [7-14]. New approaches to quantitatively identify low-abundance proteins (such as cytokines) and their functional analysis into specific protein networks are of crucial importance in the post-genomic era [15-17]. A classical proteomics approach has been used to identify in AAA subjects four serum proteins that show altered expression profiles specifically linked to AAA pathology [18,19]. Proteome analysis has been generally limited in complex samples such as serum and plasma that have a dramatically wide dynamic range of protein abundance. The use of pre-fractionated plasma samples revealed that patients with small AAA showed an increased level of the enzyme glycosyl phosphatidyl inositol-specific phospholipase D compared to controls without aneurysm [20]. However, smaller proteins may be part of larger protein complexes and hence the removal of proteins involved in complexes with high-abundance proteins such as albumin may inhibit the search for disease biomarkers [7]. In recent years, bioinformatics approaches have been applied to provide systems biology interpretations of proteomic data [21]. The focus of the present study was the application of bioinformatics to analyze all the proteins lists obtained by proteomics experiments performed on human AAA serum or plasma samples. A common approach to the visualization and examination of "omics" data involves the generation of a network of all the individual components of a given setof experiments [22,23].

Citation: Modesti A, Alberio T, Gamberi T, Fasano M, Magherini F, et al. Network Biology Analysis of the Human Abdominal Aortic Aneurysm Plasma Proteome. Austin J Proteomics Bioinform & Genomics. 2014;1(1): 10.

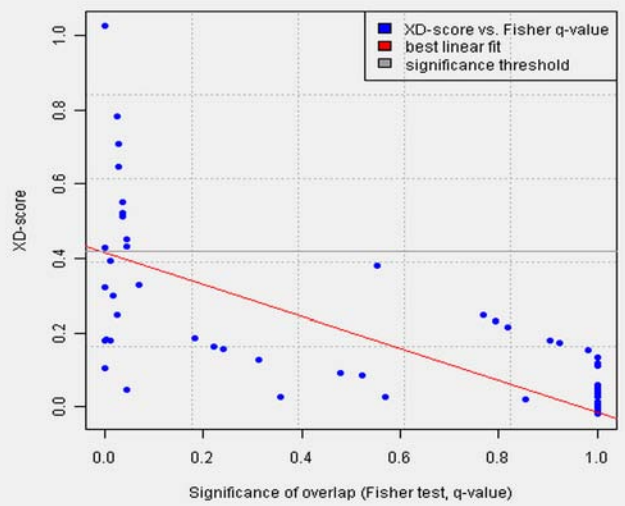

Figure 1: Linear regression analysis and Pearson correlation of Xd scores as obtained with the JEPETTO application vs the corrected p-value of the Fisher exact test. Xd values greater than the intercept of the linear model are considered as significant.

We used a network strategy to generate statistically valid hypotheses about biological mechanisms related to AAA and to provide new insight into the understanding of pathogenesis of such a complex disease [24]. To proceed we carried out a detailed bioinformatics analysis (using open source software for analyses/interpretation of biological data and visualization of complex networks obtained) on proteins identified in previous proteomics. Such an approach is being used with increased frequency, and such applications increase our understanding of complex human diseases including cardiovascular diseases [25].

Methods

Literature search to obtain the protein input list

Publications with the medical subject heading "abdominal aortic aneurysms" and keywords "proteomics" and "plasma" and "progression" or "growth" or "expansion rate" or "rupture" were searched in the MEDLINE/PubMed and EMBASE databases. The search was restricted until May 2013. Only human studies that reported aneurysm size, expansion rates and plasma markers were selected.

Bioinformatics analysis

We performed over-representation analysis (ORA) of the protein list obtained by the literature search (the input list) using the Webgestalt online tool (http://bioinfo.vanderbilt.edu/webgestalt/) against Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathway and Pathway Commons databases [26]. The analysis of functional genomics data often involves the assessment of potential functional associations between a protein set of interest. To identify these putative associations several enrichment analysis tools have been developed including over-representation analysis (ORA) techniques. This method assess the statistical over-representation of a user-defined, pre-selected gene/protein list of interest in a reference list of known gene/protein sets using a statistical test, e.g. the one-sided Fisher's exact test or the hypergeometric distribution.

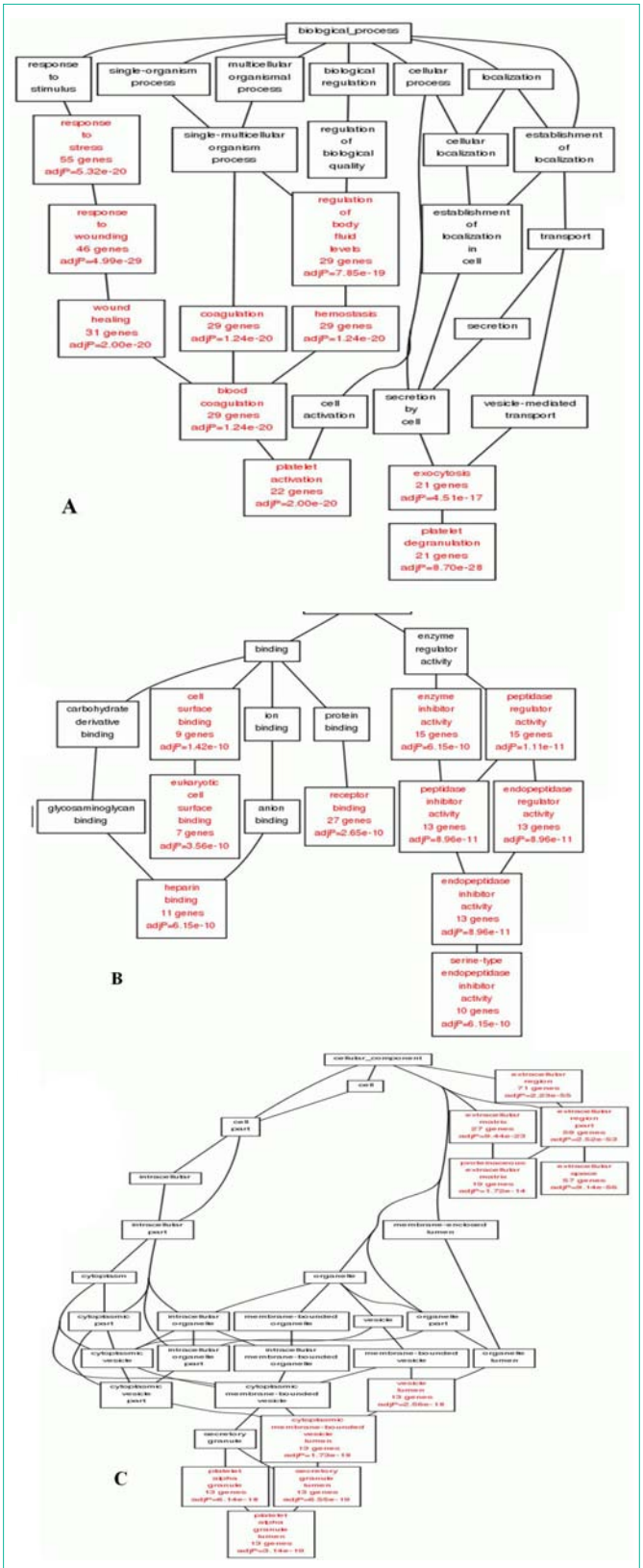

Figure 2: Gene Ontology over-representation analysis of the input list performed by Webgestalt significantly over-represented categories are highlighted in red. Panel A: GO Biological Process (BP). Panel B: GO Molecular Function (MF). Panel C: Cellular Component (CC).

Table 1: The input list.

Uniprot accession number	Gene Symbol	Protein name	Reference
O00175	CCL24	chemokine (C-C motif) ligand 24	[17]
O43927	CXCL13	chemokine (C-X-C motif) ligand 13	[17]
P00450	CP	ceruloplasmin (ferroxidase)	[14,18]
P00734	F2	coagulation factor II (thrombin)	[14,20]
P00738	HP	haptoglobin	[7,14,18]
P00748	F12	coagulation factor XII (Hageman factor)	[16]
P00751	CFB	complement factor B	[7,14]
P01009	SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	[7,14,16,20]
P01011	SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	[14]
P01023	A2M	alpha-2-macroglobulin	[7,14]
P01024	C3	complement component 3	[16]
P01042	KNG1	kininogen 1	[7]
P01579	IFNG	interferon, gamma	[20]
P01834	IGKC	immunoglobulin kappa constant	[16]
P01857	IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)	[19]
P01859	IGHG2	immunoglobulin heavy constant gamma 2 (G2m marker)	[19]
P01871	IGHM	immunoglobulin heavy constant mu	[19]
P02647	APOA1	apolipoprotein A-I	[7,14,18,20]
P02655	APOC2	apolipoprotein C-II	[13]
P02671	FGA	fibrinogen alpha chain	[7,14]
P02675	FGB	fibrinogen beta chain	[14]
P02679	FGG	fibrinogen gamma chain	[7,14,18]
P02741	CRP	C-reactive protein, pentraxin-related	[20]
P02743	APCS	amyloid P component, serum	[18]
P02750	LRG1	leucine-rich alpha-2-glycoprotein 1	[7,14]
P02753	RBP4	retinol binding protein 4, plasma	[14]
P02760	AMBP	alpha-1-microglobulin/bikunin precursor	[15]
P02765	AHSG	alpha-2-HS-glycoprotein	[7,14]
P02768	ALB	albumin	[13,14,16]
P02774	GC	group-specific component (vitamin D binding protein)	[7,18]
P02775	PPBP	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	[15]
P02787	TF	transferrin	[7,14,18]
P02790	HPX	hemopexin	[7,18]
P05546	SERPIND1	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	[15]
P06396	GSN	gelsolin	[15,19]
P06727	APOA4	apolipoprotein A-IV	[14]
P07737	PFN1	profilin 1	[15]
P08185	SERPINA6	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6	[15]
P08294	SOD3	superoxide dismutase 3, extracellular	[15]
P08567	PLEK	pleckstrin	[15]
P08697	SERPINF2	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2	[15,20]
P08833	IGFBP1	insulin-like growth factor binding protein 1	[17]
P09237	MMP7	matrix metallopeptidase 7 (matrilysin, uterine)	[20]
P0C0L5	LOC100293534	complement C4-B-like	[18]
P10909	CLU	clusterin	[7,14]
P12814	ACTN1	actinin, alpha 1	[15]
P14151	SELL	selectin L	[17]
P15169	CPN1	carboxypeptidase N, polypeptide 1	[15]
P15502	ELN	elastin	[20]
P18065	IGFBP2	insulin-like growth factor binding protein 2, 36kDa	[17]
P18206	VCL	vinculin	[15]
P19320	VCAM1	vascular cell adhesion molecule 1	[17]
P21333	FLNA	filamin A, alpha	[15]
P29622	SERPINA4	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 4	[15]

P39905	GDNF	glial cell derived neurotrophic factor	[17]
P61626	LYZ	lysozyme	[15]
P68871	HBB	hemoglobin, beta	[7,13]
P78556	CCL20	chemokine (C-C motif) ligand 20	[17]
P80108	GPLD1	glycosylphosphatidylinositol specific phospholipase D1	[19]
Q01518	CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)	[15]
Q14624	ITIH4	inter-alpha-trypsin inhibitor heavy chain family, member 4	[19]
Q15848	ADIPOQ	adiponectin, C1Q and collagen domain containing	[15,17]
Q96IY4	CPB2	carboxypeptidase B2 (plasma)	[15]
Q9Y490	TLN1	talin 1	[15]

Afterwards the web portal BioProfiling.de (http://www.bioprofiling.de/) was used to build enriched networks. This portal covered most of the available information regarding signaling and metabolic pathways (database: Reactome) and physical protein interactions (database: IntAct) [27]. Accordingly, the input list was analyzed by Rspider [28] and PPI spider [29,30]. Both tools employed advanced enrichment or network-based statistical frameworks. The p-value provided was computed by the Monte Carlo simulation (http://www.bioprofiling.de/statistical_frameworks.html) and referred to the probability of obtaining a model of the same quality for a random gene list of the same size [31]. The significant networks, using a cut-off of $p < 0.01$, were further considered for proteomics data interpretation and discussion. The enriched network was exported as an .xgmml file and visualized and modified by Cytoscape 3.1.0 (http://www.cytoscape.org/) [32]. Moreover, the Cytoscape Jepetto (Java Enrichment of Pathways Extended to Topology) application was employed to perform ORA of the enriched networks [33] (http://apps.cytoscape.org/apps/jepetto). For the enrichment analysis by JEPETTO, EnrichNet server is used. EnrichNet maps the input gene set onto a molecular interaction network and, using a random walk, scores distances between the genes and pathways/processes in a reference database. This network-based association score (XD-score) is relative to the average distance to all pathways and represents a deviation (positive or negative) from the average distance. Significantly over-represented pathways were selected if theirXd score was greater than that calculated as the intercept of the linear regression of Xd scores vs the Fisher's exact test p-value after Benjamini-Hochberg correction (q-value; Figure 1) [34].

Results and Discussion

Literature analysis of proteomics studies of plasma AAA associated markers. Nine published proteomics studies were used to generate an input protein set of 64 seed proteins (the input list). The proteins included in the input list were differentially expressed in plasma from AAA patients with respect to controls. The 64 seed proteins are reported in Table 1.

Over-representation analysis of the input list

The input list was uploaded to the Webgestalt portal and ORA was performed. Figure 2 shows the significantly over-represented GO Biological Processes (BP, Panel A), Molecular Functions (MF, Panel B) and Cellular Components (CC, Panel C) in the input protein set (for details, see Supplementary Table 1). As expected for this pathology, we found enriched Gene Ontology Biological Process (GO BP) terms involved in response mechanisms to generic stress conditions, including response to stress (GO:0006950 - 47 genes); and inflammatory response (GO:0006954 – 20 genes). Other BP terms appear to be more closely related to the pathological condition being investigated, namely hemostasis (GO: 0007599 – 25 genes); blood coagulation (GO: 0007596 – 25 genes); and the related protein activation cascade (GO: 0072376 – 8 genes). Of note, some of the BP terms corresponded to processes involved in transport and secretion,

Table 2: Over-representation of KEGG pathways in the input dataset. $p < 0.001$ after Benjamini-Hochberg correction.

PathwayName	#Gene	EntrezGene	Statistics
Complement and coagulation cascades	14	P0C0L5 P01024 P01009 Q96IY4 P05546 P01023 P01042 P02675 P02679 P00751 P02671 Q8IZZ5 P00748	C=69;O=14;E=0.10;R=136.73;rawP=7.91e-27;adjP=1.66e-25
African trypanosomiasis	4	P01579 P68871 P19320 P02647	C=35;O=4;E=0.05;R=77.01;rawP=2.23e-07;adjP=2.34e-06
Staphylococcus aureus infection	4	P0C0L5 P01024 P02679 P00751	C=55;O=4;E=0.08;R=49.01;rawP=1.42e-06;adjP=9.94e-06
Regulation of actin cytoskeleton	5	P06396 P12814 P00734 P18206 P07737	C=213;O=5;E=0.32;R=15.82;rawP=1.69e-05;adjP=8.87e-05
Systemic lupus erythematosus	4	P0C0L4 P01579 P01024 P12814	C=136;O=4;E=0.20;R=19.82;rawP=5.19e-05;adjP=0.0002
Cytokine-cytokine receptor interaction	5	P01579 O00175 O43927 P78556 P02775	C=265;O=5;E=0.39;R=12.71;rawP=4.78e-05;adjP=0.0002
Malaria	3	P01579 P68871 P19320	C=51;O=3;E=0.08;R=39.64;rawP=6.17e-05;adjP=0.0002
Focal adhesion	4	P12814 P21333 P18206 Q9Y490	C=200;O=4;E=0.30;R=13.48;rawP=0.0002;adjP=0.0005
Chemokine signaling pathway	4	O00175 O43927 P78556 P02775	C=189;O=4;E=0.28;R=14.26;rawP=0.0002;adjP=0.0005
Amoebiasis	3	P01579 P12814 P18206	C=106;O=3;E=0.16;R=19.07;rawP=0.0005;adjP=0.0010

The statistics column lists: C: the number of reference genes in the category; O: the number of genes in the gene set and also in the category; E: the expected number in the category; R: ratio of enrichment; rawP: p value from the Fisher exact test; adjP: p value adjusted for multiple testing.

Table 3: Over - Representation of pathway commons pathways in the input dataset. p < 0.001 after Benjamini-Hochberg correction.

PathwayName	#Gene	EntrezGene	Statistics
Hemostasis	18	P00734 Q9Y490 P01023 P14151 P01042 P02679 P00748 P21333 P02787 P68871 P08567 Q01518 P07737 P02675 P02671 P08697 P02647 P18206	C=376;O=18;E=0.56;R=32.26;rawP=1.41e-22;adjP=1.71e-20
Platelet degranulation	8	Q9Y490 P21333 P02787 P08567 Q01518 P07737 P02647 P18206	C=29;O=8;E=0.04;R=185.89;rawP=6.25e-17;adjP=3.78e-15
Proteoglycan syndecan-mediated signaling events	21	P01579 P00734 P00450 Q9Y490 P08833 P01023 P25090 P02679 P12814 P21333 P02787 P80108 P02765 P06396 P09237 P39905 P02675 P02671 P10909 P18206 P02741	C=1345;O=21;E=2.00;R=10.52;rawP=2.32e-16;adjP=9.36e-15
Integrin family cell surface interactions	21	P01024 P01579 P00450 Q9Y490 P08833 P01023 P25090 P02679P12814 P02787 P80108 P02765 P06396 P09237 P39905 P19320 P02675 P02671 P10909 P18206 P02741	C=1378;O=21;E=2.04;R=10.27;rawP=3.76e-16;adjP=1.14e-14
Response to elevated platelet cytosolic Ca2+	8	Q9Y490 P21333 P02787 P08567 Q01518 P07737P02647P18206	C=40;O=8;E=0.06;R=134.77;rawP=1.11e-15;adjP=2.69e-14
Alpha9 beta1 integrin signaling events	19	P01579 P00450 Q9Y490 P08833 P01023 P02679 P12814 P02787 P80108 P02765 P06396 P09237 P19320 P39905 P02675 P02671 P10909 P18206 P02741	C=1305;O=19;E=1.94;R=9.81;rawP=2.89e-14;adjP=4.37e-13
PAR1-mediated thrombin signaling events	19	P01579 P00734 P00450 Q9Y490 P08833 P01023 P02679 P12814 P02787 P80108 P02765 P06396 P09237 P39905 P02675 P02671 P10909 P18206 P02741	C=1299;O=19;E=1.93;R=9.86;rawP=2.66e-14;adjP=4.37e-13
Thrombin/protease-activated receptor (PAR) pathway	19	P01579 P00734 P00450 Q9Y490 P08833 P01023 P02679 P12814 P02787 P80108 P02765 P06396 P09237 P39905 P02675 P02671 P10909 P18206 P02741	C=1300;O=19;E=1.93;R=9.85;rawP=2.69e-14;adjP=4.37e-13
Formation of Fibrin Clot (Clotting Cascade)	7	P00734 P01023 P01042 P02679 P00748 P02675 P02671	C=33;O=7;E=0.05;R=142.94;rawP=4.68e-14;adjP=5.66e-13

The statistics column lists: C: the number of reference genes in the category; O: the number of genes in the gene set and also in the category; E: the expected number in the category; R: ratio of enrichment; rawP: p value from the Fisher exact test; adjP: p value adjusted for multiple testing.

such as secretion by cell (GO: 0032940 – 24 genes); exocytosis (GO: 0006887 – 19 genes); and platelet degranulation (GO: 0002576 – 19 genes). As expected, the molecular functions related to coagulation proteases, including regulatory factors, were significantly enriched in the input list. Molecular function terms associated with receptor binding were over-represented in the list. The enriched GO MF terms contain ontologies involved in binding, including cell surface binding (GO: 0043498 – 7 genes); protein binding (GO: 0005515 - 47 genes); and receptor binding (GO: 0005102 – 19 genes). There were also enriched GO MF terms involved in enzyme regulator activity, including enzyme inhibitor (GO: 0004857 - 13 genes); peptidase regulator (GO: 0061134 - 12 genes); peptidase inhibitor (GO:0030414-13 genes); endopeptidase regulator (GO:0061135-12 genes) and endopeptidase inhibitor (GO:0004866 - 12 genes). The enriched GO cellular components terms pointed out two major over-represented classes. The first were associated with the secretory pathway, which would be expected for platelet activation and included platelet alpha granule lumen (GO: 0031093 - 11 genes); and platelet alpha granule (GO: 0031091 – 11 genes). The other class was those components on the cell surface or in the extracellular region and included extracellular space (GO: 0005615 – 41 genes); extracellular region (GO: 0005576 – 54 genes); extracellular region part (GO:0044421 – 42 genes); and extracellular matrix (GO:0031012 – 10 genes), suggesting a possible role of cell-celland cell-ECM adhesion in the composition of the investigated signature. The set was then analyzed with respect to: KEGG (Table 2), Pathway Commons (Table 3) and Wiki Pathways (Table 4) databases. A significant over-representation of pathways related to blood coagulation and thrombin/plasmin cascades was found (14 genes), as expected. Additionally, evidence for a remarkable involvement of cell-mediated responses came to the light. Indeed, GO ORA suggested the role of soluble platelet activation markers as detectable in plasma. KEGG pathways indicated that a few proteins in the list were linked to focal adhesion (4 genes), cytokine-cytokine receptor interaction (5 genes) and chemokine signaling (4 genes). Results from the PathwayCommons ORA also supported the link of the proteins set with either cell-cell or cell-ECM adhesion processes (integrin family cell surface interaction - 21 genes, alpha9 beta1 integrin signaling events - 19 gens, proteoglycan syndecan-mediated signaling events - 21 genes), in addition to pathways associated with blood coagulation and fibrinolysis (hemostasis - 18 genes; thrombin/protease-activated receptor (PAR) pathway and PAR1-8 mediated thrombin signaling events - 19 genes for each pathway). In addition, WikiPathways ORA indicated a possible connection between the input list and reactive oxygen species scavenging (selenium pathway - 8 genes) and biosynthesis. As a whole, the ORA performed on the input list identified several elements of the protein signature not

Table 4: Over-Representation of WikiPathways entries in the input dataset. p < 0.001 after Benjamini-Hochberg correction.

PathwayName	#Gene	EntrezGene	Statistics
Complement and Coagulation Cascades	10	P01024 P00734 P01009 P05546 Q96IY4 P01023 P01042 P02675 P00748 P08697	C=51;O=10;E=0.08;R=132.13;rawP=3.01e-19;adjP=4.52e-18
Blood Clotting Cascade	6	P00734 P02679 P02675 P00748 P02671 P08697	C=26;O=6;E=0.04;R=155.51;rawP=1.89e-12;adjP=1.42e-11
Selenium Pathway	8	P68871 P01579 P00734 P02768 P01011 P02647 P02741 P08294	C=108;O=8;E=0.16;R=49.92;rawP=4.68e-12;adjP=2.34e-11
Regulation of Actin Cytoskeleton	5	P00734 P06396 P07737 P12814 P18206	C=157;O=5;E=0.23;R=21.46;rawP=3.85e-06;adjP=1.44e-05
Vitamin B12 Metabolism	3	P68871 P08294 P02647	C=24;O=3;E=0.04;R=84.23;rawP=6.17e-06;adjP=1.85e-05
Folate Metabolism	3	P68871 P08294 P02647	C=29;O=3;E=0.04;R=69.71;rawP=1.11e-05;adjP=2.64e-05
Statin Pathway	3	P02655 P06727 P02647	C=30;O=3;E=0.04;R=67.39;rawP=1.23e-05;adjP=2.64e-05

The statistics column lists: C: the number of reference genes in the category; O: the number of genes in the gene set and also in the category; E: the expected number in the category; R: ratio of enrichment; rawP: p value from the Fisher exact test; adjP: p value adjusted for multiple testing.

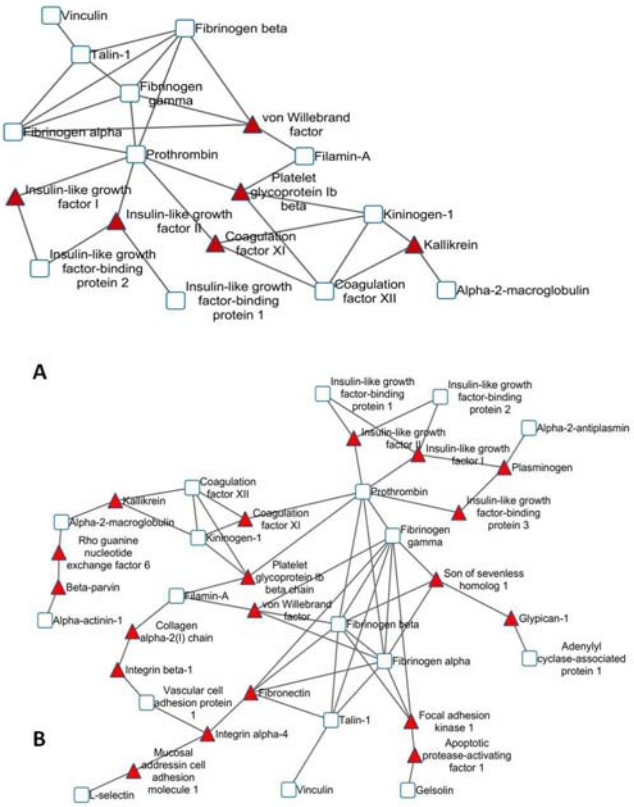

Figure 3: Enriched networks obtained with Rspider. Panel A: A maximum of one intermediate node was allowed to be added. Panel B**:** A maximum of two intermediate nodes was allowed to be added. Network visualization was obtained using Cytoscape. Nodes arising from the "input list" are shown as white squares, whereas enriched nodes are represented by red triangles. p < 0.01.

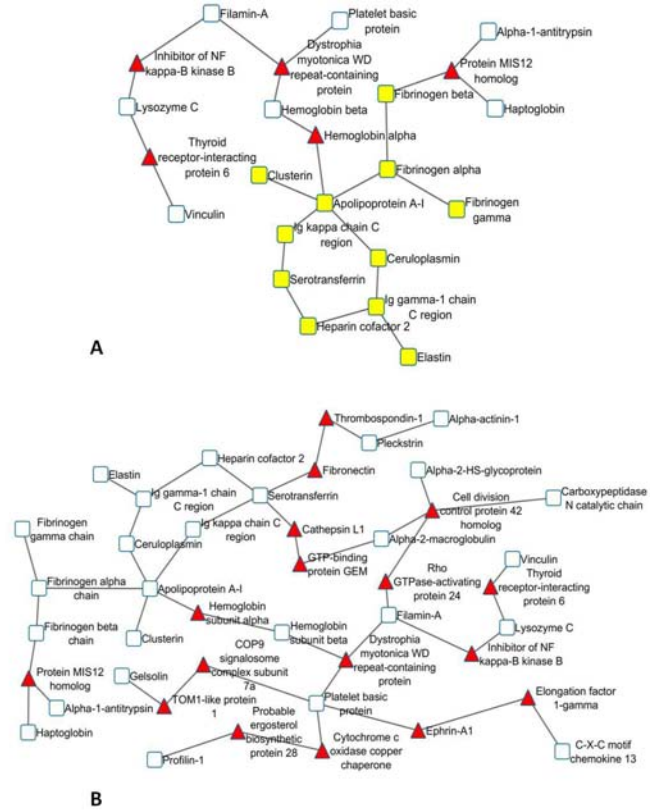

Figure 4: Enriched networks obtained with PPI spider. Panel A: A maximum of one intermediate node was allowed to be added. Panel B**:** A maximum of two intermediate nodes was allowed to be added. Network visualization was obtained using Cytoscape. p < 0.01. Nodes arising from the "input list" are shown as white/yellow squares, whereas enriched nodes are represented by red triangles. A highly-interconnected sub network without added nodes is highlighted in yellow.

directly related to wound healing and/or coagulation. Among those, the presence of soluble platelet activation markers and of proteins involved in cellular processes at the cell surface could represent an interesting hint to look for novel prognostic AAA markers. To further reinforce this finding, we performed network enrichment analysis.

Network enrichment

In order to discriminate the number of proteins that could be associated to disease, the protein set was enriched in terms of a network-based statistical framework. As a first approach, a pathway-based network was built by allowing the insertion of no more than 1 or 2 intermediate nodes using the Reactome database as a reference (Figure 3). In both cases, highly interconnected networks were obtained, with 12 nodes (Figure 3, Panel A) and 18 nodes (Figure 3, Panel B) overlapping with the input list, respectively. Furthermore, a physical-interaction-based network was built by limiting the insertion of enriched nodes to one or two using the Int Act database

Table 5: Over-Representation analysis of the network displayed in Figure 3, Panel A.

GO Category (Biological Process)	XD-score	q-value	Overlap/Size
blood coagulation, intrinsic pathway	1,63249	0,00000	7/17
regulation of blood coagulation	1,18543	0,00021	3/10
positive regulation of blood coagulation	0,78543	0,02288	2/10
positive regulation of activated T cell proliferation	0,78543	0,00065	3/15
positive regulation of glycogen biosynthetic process	0,71270	0,02540	2/11
fibrinolysis	0,65210	0,00099	3/18
protein localization at cell surface	0,55686	0,03839	2/14
GO Category (Molecular Function)	**XD-score**	**q-value**	**Overlap/Size**
insulin-like growth factor receptor binding	0,65619	0,01662	2/12
eukaryotic cell surface binding	0,51126	0,00296	3/23
insulin-like growth factor binding	0,46011	0,02919	2/17
insulin receptor binding	0,40332	0,00298	3/29
Rho GTPase binding	0,33735	0,04721	2/23
GO Category (Cellular Component)	**XD-score**	**q-value**	**Overlap/Size**
platelet alpha granule	1,33158	0,00000	4/12
platelet alpha granule lumen	0,59768	0,00000	7/46

Table 6: Over-representation analysis of the network displayed in Figure 3, Panel B.

GO Category (Biological Process)	XD-score	q-value	Overlap/Size
blood coagulation, intrinsic pathway	1,61553	0,00000	7/17
negative regulation of fibrinolysis	1,16847	0,00107	3/10
regulation of blood coagulation	1,16847	0,00107	3/10
fibrinolysis	1,07959	0,00000	5/18
positive regulation of blood coagulation	0,76847	0,04487	2/10
positive regulation of activated T cell proliferation	0,76847	0,00334	3/15
positive regulation of insulin-like growth factor receptor signaling pathway	0,76847	0,04487	2/10
platelet degranulation	0,75795	0,00000	15/76
positive regulation of glycogen biosynthetic process	0,69575	0,05110	2/11
tissue regeneration	0,67436	0,00459	3/17
protein localization at cell surface	0,53990	0,06812	2/14
positive regulation of collagen biosynthetic process	0,53990	0,06812	2/14
myoblast differentiation	0,53990	0,06812	2/14
leukocyte cell-cell adhesion	0,49021	0,01018	3/23
protein polymerization	0,49021	0,01018	3/23
GO Category (Molecular Function)	**XD-score**	**q-value**	**Overlap/Size**
insulin-like growth factor binding	0,68349	0,00260	3/17
eukaryotic cell surface binding	0,67326	0,00040	4/23
fibronectin binding	0,64427	0,00260	3/18
insulin-like growth factor receptor binding	0,64427	0,03254	2/12
laminin binding	0,51094	0,04795	2/15
protease binding	0,40550	0,00040	5/47
insulin receptor binding	0,39140	0,00902	3/29
GO Category (Cellular Component)	**XD-score**	**q-value**	**Overlap/Size**
platelet alpha granule	1,31763	0,00001	4/12
platelet alpha granule lumen	0,93156	0,00000	11/46
fascia adherens	0,59042	0,03764	2/13
cortical actin cytoskeleton	0,39609	0,07391	2/19
podosome	0,37504	0,53453	1/10

Table 7: Over-Representation analysis of the subnetwork highlighted in Figure 4, Panel A.

GO Category (Biological Process)	XD-score	q-value	Overlap/Size
reverse cholesterol transport	0,72179	0,01239	2/11
protein polymerization	0,51626	0,00047	3/23
triglyceride homeostasis	0,39452	1,00000	1/10
endothelial cell proliferation	0,39452	1,00000	1/10
complement activation	0,35815	0,04435	2/22
cholesterol efflux	0,35815	1,00000	1/11
lipid storage	0,32785	1,00000	1/12
iron ion transport	0,32785	1,00000	1/12
platelet degranulation	0,31031	0,00000	6/76
GO Category (Molecular Function)	**XD-score**	**q-value**	**Overlap/Size**
eukaryotic cell surface binding	0,51797	0,00055	3/23
lipid transporter activity	0,32956	0,91340	1/12
GO Category (Cellular Component)	**XD-score**	**q-value**	**Overlap/Size**
platelet alpha granule	0,99286	0,00002	3/12
platelet alpha granule lumen	0,34068	0,00001	4/46
very-low-density lipoprotein particle	0,32619	0,43292	1/12

Table 8: Over-representation analysis of the network shown in Figure 4, Panel A.

GO Category (Biological Process)	XD-score	q-value	Overlap/Size
reverse cholesterol transport	0,71498	0,05011	2/11
platelet degranulation	0,57981	0,00000	10/76
protein localization at cell surface	0,55914	0,07227	2/14
protein polymerization	0,50945	0,00533	3/23
GO Category (Molecular Function)	**XD-score**	**q-value**	**Overlap/Size**
eukaryotic cell surface binding	0,51256	0,00523	3/23
GO Category (Cellular Component)	**XD-score**	**q-value**	**Overlap/Size**
platelet alpha granule	0,98688	0,00013	3/12
platelet alpha granule lumen	0,61731	0,00000	6/46
Wiki Pathways	**XD-score**	**q-value**	**Overlap/Size**
Blood Clotting Cascade	0,55235	0,00472	3/21
Reactome	**XD-score**	**q-value**	**Overlap/Size**
COMMON PATHWAY	0,89808	0,00091	3/13
GRB2 SOS PROVIDES LINKAGE TO MAPK SIGNALING FOR INTEGRINS	0,77500	0,00108	3/15
P130CAS LINKAGE TO MAPK SIGNALING FOR INTEGRINS	0,77500	0,00108	3/15
FURTHER PLATELET RELEASATE	0,73691	0,00007	4/21
PLATELET DEGRANULATION	0,53750	0,00000	10/80
KEGG	**XD-score**	**q-value**	**Overlap/Size**
Complement and coagulation cascades	0,29122	0,00041	5/65

as a reference (Figure 4). In the resulting networks, 18 nodes (Figure 4, Panel A) and 26 nodes (Figure 4, Panel B) overlapped with the input list, respectively. Interestingly, a subnetwork of 11 interacting proteins was detected without any added node. References for the protein-protein interaction pairs defining the edges of Figure 4 are reported in the Supplementary Table 2.

Over-representation analysis of the enriched networks

Enriched networks were analyzed by ORA within the specific network-based environment provided by the Cytoscape platform. As far as pathway-based networks are concerned, the analysis was limited to GO terms. As a first approach, the protein set of Figure 3A was analyzed, and the results are listed in Table 5. In addition to GO categories intrinsically linked to blood coagulation, a BP term associated with T-cell involvement was significantly over-represented, suggesting a role of lymphocyte cell-mediated responses in the composition of the plasma proteome AAA fingerprint. The same is supported by MF terms being over-represented in the present analysis, such as signal transduction pathways associated to surface binding and growth factor receptors. However, the analysis of this enriched network did not add a significant amount of new information to the analysis of the input list. Therefore, a more extended enrichment was taken into account. The analysis of the extended network (Figure 3, Panel B) revealed the identification of other categories possibly involved in the AAA pathogenetic process (Table 6). By allowing a maximum of two intermediate nodes to be inserted in the pathway-based network, stronger roles for cell-cell/ECM interaction and leukocyte-mediated response became appreciable in the plasma proteome. The ORA procedure was also performed on the protein-protein interaction networks. A physical-interaction-based enrichment procedure may potentially explore distinct pathways and/or GO categories that would have been restricted by a pathway-based procedure. Accordingly, the highly interconnected subnetwork highlighted in Figure 4A was analyzed as

a standalone network (Table 7), and the whole graph was later taken into account to find over-represented GO categories and reactome pathways (Table 8). The analysis of the interconnected subcluster highlighted in Figure 4A drew attention to lipid storage and transport and again an involvement for platelet degranulation. The analysis of the whole network indicated the potential molecular mediators of the cellular response evidenced above. In particular, the signal transduction pathways linked to integrin signaling such as the GRB2/SOS/ERK cascade was highly prevalent. The same ORA procedure applied to the network reported in Figure 4B did not add any other significant category/pathway to the results (Supplementary Table 3).

Concluding remarks

A thorough systems biology analysis of the complex plasma protein panel associated with AAA has highlighted the relevant presence of proteins linked to blood coagulation and to the proteolytic enzyme cascade leading to blood clotting. Moreover, input list of proteins uncovered significant functional connections with biological processes and molecular functions mediated by blood cells. In particular, cell adhesion mechanisms mediated by integrins, signal transduction pathways elicited by cytokines and growth factors, and markers of the inflammatory response represent a valuable pool of candidate novel biomarkers of AAA outcomes. As reported by Albert-László Barabási [22] an integrated understanding of the interactions among the genome, the proteome, the environment and the pathophenome, mediated by the underlying cellular network, offers a basis for the cells' global organization - the 'think globally, act locally' paradigm of network medicine.

References

1. Sakalihasan N, Limet R, Defawe OD Abdominal aortic aneurysm. Lancet 2005; 365: 1577 -1589.
2. Lederle FA, Kane RL, MacDonald R, Wilt TJ. Systematic review: repair of unruptured abdominal aortic aneurysm. Ann Intern Med 2007; 146: 735-741.
3. Thompson SG, Ashton HA, Gao L, Scott RA. Screening men forminal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomized Multicentre Aneurysm Screening Study. BMJ 2009; 338: b2307.
4. Lederle FA. Vascular disease: is AAA screening worth the cost? Nat Rev Cardiol 2009; 6: 616-618.
5. Baxter BT, Terrin MC, Dalman RL. Medical management of small abdominal aortic aneurysms. Circulation 2008; 117: 1883-1889.
6. Golledge J, Tsao PS, Dalman RL, Norman PE. Circulating markers of abdominal aortic aneurysm presence and progression. Circulation 2008; 118: 2382–2392.
7. Gamberi T, Puglia M, Guidi F, Magherini F, Bini L, et al. A proteomic approach to identify plasma proteins in patients with abdominal aortic aneurysm. Mol Biosyst 2011; 7: 2855-2862.
8. Urbonavicius S, Urbonaviciene G, Honoré B, Henneberg EW, Vorum H, et al. Potential circulating biomarkers for abdominal aortic aneurysm expansion and rupture--a systematic review. Eur J Vasc Endovasc Surg 2008; 36: 273-280.
9. Lindholt JS, Ashton HA, Heickendorff L, Scott RA. Serum elastin peptides in the preoperative evaluation of abdominal aortic aneurysms. Eur J Vasc Endo vasc Surg 2001; 22: 546-550.
10. Lindholt JS, Jørgensen B, Fasting H, Henneberg EW. Plasma levels of plasmin-antiplasmin-complexes are predictive for small abdominal aortic aneurysms expanding to operation-recommendable sizes. J Vasc Surg 2001; 34: 611-615.
11. Yamazumi K1, Ojiro M, Okumura H, Aikou T. An activated state of blood coagulation and fibrinolysis in patients with abdominal aortic aneurysm. Am J Surg 1998; 175: 297-301.
12. Al-Barjas HS, Ariëns R, Grant P, Scott JA. Raised plasma fibrinogen concentration in patients with abdominal aortic aneurysm. Angiology 2006; 57: 607-614.
13. Nordon IM, Brar R, Hinchliffe RJ, Cockerill G, Thompson MM. Proteomics and pitfalls in the search for potential biomarkers of abdominal aortic aneurysms. Vascular 2010; 18: 264-268.
14. Modesti PA, Gamberi T, Bazzini C, Borro M, Romano SM, et al. Response of serum proteome in patients undergoing infrarenal aortic aneurysm repair. Anesthesiology 2009; 111: 844-854.
15. Acosta-Martin AE, Panchaud A, Chwastyniak M, Dupont A, Juthier F,et al. Quantitative mass spectrometry analysis using PAcIFIC for the identification of plasma diagnostic biomarkers for abdominal aortic aneurysm. PLoS One 2011; 6: 28698.
16. Pulinx B, Hellenthal FA, Hamulyák K, van Dieijen-Visser MP, Schurink GW, et al. Differential protein expression in serum of abdominal aortic aneurysm patients - a proteomic approach. Eur J Vasc Endovasc Surg 2011; 42: 563-570.
17. Ramos-Mozo P, Rodriguez C, Pastor-Vargas C, Blanco-Colio LM, Martinez-Gonzalez J, et al. Plasma profiling by a protein array approach identifies IGFBP-1 as a novel biomarker of abdominal aortic aneurysm. Atherosclerosis 2012; 221: 544-550.
18. Spadaccio C, Di Domenico F, Perluigi M, Lusini M, Giorgi A, et al. Serum proteomics in patients with diagnosis of abdominal aortic aneurysm. Cardio vasc Pathol 2012; 21: 283-290.
19. Wallinder J, Bergström J, Henriksson AE. Discovery of a novel circulating biomarker in patients with abdominal aortic aneurysm: a pilot study using a proteomic approach. ClinTransl Sci 2012; 5: 56-59.
20. Ehsan S, Ball G, Choke E, Molyneux KM, London NJ, et al. Disease specific biomarkers of abdominal aortic aneurysms detected by surface enhanced laser desorption ionization time of flight mass spectrometry. Eur J Vasc Endovasc Surg 2012; 44: 52-54.
21. Gamberi T, Magherini F, Bini L, Messori L. New Insights into the Molecular Mechanisms of Selected Anticancer Metal Compounds through Bioinformatic Analysis of Proteomic Data. J. Proteomics Bioinform 2013, S6: 006.
22. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011; 12: 56-68.
23. Diez D, Wheelock AM, Goto S, Haeggström JZ, Paulsson-Berne G, et al. The use of network analyses for elucidating mechanisms in cardiovascular disease. Mol Biosyst 2010; 6: 289-304.
24. Chan SY, Loscalzo J. The emerging paradigm of network medicine in the study of human disease. Circ Res 2012; 111: 359-374.
25. Chan SY, White K, Loscalzo J. Deciphering the molecular basis of human cardiovascular disease through network biology. Curr Opin Cardiol 2012; 27: 202-209.
26. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis. Toolkit (WebGestalt): update 2013. Nucleic Acids Res 2013; 41: W77-83.
27. Antonov AV. BioProfiling.de: analytical web portal for high-throughput cell biology. Nucleic Acids Res 2011; 39: W323-W327.
28. Haw R, Stein L. Using the Reactome database. CurrProtoc Bioinformatics 2012; 38:8.7:8.7.1–8.7.23.
29. Antonov AV, Dietmann S, Rodchenkov I, Mewes HW. PPI spider: a tool for the interpretation of proteomics data in the context of protein-protein interaction networks. Proteomics 2009; 9: 2740-2749.
30. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 2014; 42: 358-363.
31. Antonov AV, Schmidt E, Dietmann S, Krestyaninova M, Hermjakob H. R spider: a network-based analysis of gene lists by combining signaling and

metabolic pathways from Reactome and KEGG databases. Nucleic Acids Research 2010; 38: W78-83.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.

33. Winterhalter C, Widera P, Krasnogor N, JEPETTO: a Cytoscape plugin for gene set enrichment and topological analysis based on interaction networks. Bioinformatics 2014; 30: 1029-1030.

34. Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. Enrich Net: network-based gene set enrichment analysis. Bioinformatics 2012; 28: 451-457.

Citation: Modesti A, Alberio T, Gamberi T, Fasano M, Magherini F, et al. Network Biology Analysis of the Human Abdominal Aortic Aneurysm Plasma Proteome. Austin J Proteomics Bioinform & Genomics. 2014;1(1): 10.