

RESEARCH ARTICLE

Open Access

Preeclampsia: a bioinformatics approach through protein-protein interaction networks analysis

Eduardo Tejera¹, João Bernardes² and Irene Rebelo^{1*}

Abstract

Background: In this study we explored preeclampsia through a bioinformatics approach. We create a comprehensive genes/proteins dataset by the analysis of both public proteomic data and text mining of public scientific literature. From this dataset the associated protein-protein interaction network has been obtained. Several indexes of centrality have been explored for hubs detection as well as the enrichment statistical analysis of metabolic pathway and disease.

Results: We confirmed the well known relationship between preeclampsia and cardiovascular diseases but also identified statistically significant relationships with respect to cancer and aging. Moreover, significant metabolic pathways such as apoptosis, cancer and cytokine-cytokine receptor interaction have also been identified by enrichment analysis. We obtained FLT1, VEGFA, FN1, F2 and PGF genes with the highest scores by hubs analysis; however, we also found other genes as PDIA3, LYN, SH2B2 and NDRG1 with high scores.

Conclusions: The applied methodology not only led to the identification of well known genes related to preeclampsia but also to propose new candidates poorly explored or completely unknown in the pathogenesis of preeclampsia, which eventually need to be validated experimentally. Moreover, new possible connections were detected between preeclampsia and other diseases that could open new areas of research. More must be done in this area to resolve the identification of unknown interactions of proteins/genes and also for a better integration of metabolic pathways and diseases.

Background

Preeclampsia is a pregnancy related disease associated with hypertension, proteinuria and increased maternal and perinatal morbidity and mortality, without known underlying mechanism and preventive treatment [1,2]. On the other hand, the future health or possible risks of women with previous history of preeclampsia are important areas of investigation. In this direction, it is well known the increased risk of future cardiovascular disease and renal dysfunction, however, other risks are also being discussed [1,3-5]. Owing to the morbidity and mortality of this pregnancy related disease and the possible multifactorial causes involved [1-5], several experimental procedures have been applied by researchers in the last two decades, evidently, generating an elevated number of unprocessed information.

Although some bioinformatic analysis has been performed in particular microarray assays [6,7], an extensive data evaluation and processing has not yet been performed. Furthermore, the capabilities of bioinformatics tools for gene prioritization, network analysis, gene ontology and gene-disease relationships [8,9], together with all available data on protein/gene expression during preeclampsia bring an interesting and valuable opportunity for an *in-silico* study of the disease. Therefore, the present study is focused on two main areas: I) collection and basic analysis of the genes/proteins-diseases dataset, including, protein-protein interaction network and pathway enrichment analysis and II) exploration of the related gene-diseases in order to evaluate other genetic diseases possibly related with preeclampsia.

Results

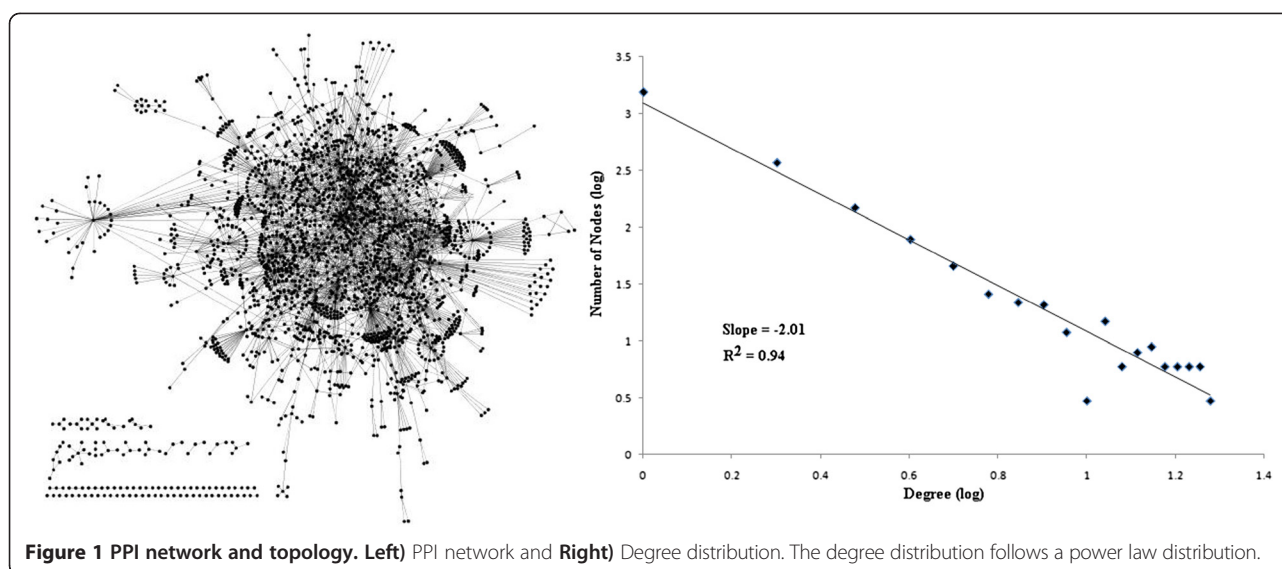
Protein-protein interaction network analysis

Preeclampsia PPI network topology reveals (Figure 1) a similar behavior with respect to general topology of PPI

* Correspondence: irebelo@ff.up.pt

¹Department of Biological Sciences, Biochemistry, Faculty of Pharmacy, University of Porto, Portugal/Institute for Molecular and Cell Biology (IBMC), Porto, Portugal

Full list of author information is available at the end of the article



following a power law behavior [10] and therefore scale-free properties. These types of networks have the particular feature that some nodes are highly connected compared with others on the same network. These highly connected nodes (hubs) in general, represent important proteins/genes in biological terms and therefore are treated with special attention.

The top 50 genes with high scores and also present in the initial set (347) are shown in Table 1, however, other genes were found with high scores value but there are not part of the initial gene group. As expected some of the selected genes like FN1, FLT1, F2, VEGFA, PGF, TNF, NOS and INHBA, are well known preeclampsia relates genes (see discussion) and several of them are related with signaling pathways.

A total of 27 communities ($k=3$) covering 161 genes were identified by communality analysis. In Figure 2 (Left) we represent those communities that are superimposed to form a large connected graph. The genes involved in communities overlapping are also highly represented in the Table 1 (and also the genes members of the large community). The model based clustering analysis reveal an optimal number of 8 clusters ($BIC = 152192.4$) with an ellipsoidal distribution with equal volume, shape and variable orientation. The genes are grouped in the clusters (C1...8) as follow: C1 (67), C2 (56), C3 (1806), C4 (59), C5 (133), C6 (23), C7 (95) and C8 (161). The C8 and C4 correspond with the highest mean scoring value: 393.3 and 348.9 respectively, and contain all the 100 genes with highest score values (part or not of the initial gene set). Furthermore, 161 genes of C8 are also the same genes detected in the communality analysis.

Gene ontology (GO) enrichment analyses were performed in all obtained clusters. However, for simplicity only C4 and C8 are presented (Figure 2 Right). The GO

analysis reveals that C8 comprise several processes related with angiogenesis, apoptosis and cell proliferation and also shared with C4 several processes involved in cell activation and biological adhesion. The relation between these processes as well as the fact that both groups are representative of the highest scored genes could indicate a particular relevance of the clusters in terms of genes-disease relationship. On the other hand, also these processes are well known involved in preeclampsia and are also consistent with the pathway enrichment analysis.

Diseases and metabolic pathway enrichment analysis

Several types of diseases were found statistically significant in the enrichment analysis; partial results are presented in Table 2. Obviously, preeclampsia and even hypertension have to be present in the analysis. In the GAD database there are several disease classes and beside the presented in Table 2, others like hematologic ($p\text{-value} = 6.1E-06$) and renal diseases ($p\text{-value} = 2.5E-04$) were also significant. Even when we present only the results obtained with GAD database, analysis was also performed with OMIN database confirming the ovarian cancer ($p\text{-value} = 0.019$) and also indicating colorectal cancer as statistically significant ($p\text{-value} = 0.011$), however, better results were obtained with GAD and also more consistent with literature information and pathway enrichment analysis.

It is important to consider that several genes in the PPI network do not present a known relation with specific diseases, at least reported in the GAD or OMIN databases. Only around 30% of the 2400 genes were found in the databases. This difficulty means that we have to be cautious with the preeclampsia genes-diseases relationships and with the reliability of the

Table 1 Top 50 genes obtained by analysis of the PPI network

| Gene symbol | Entrez ¹ | Score I | Gene symbol | Entrez ¹ | Score I |
|-------------|---------------------|----------|-------------|---------------------|----------|
| 4067 | LYN | 983.5136 | 2324 | FLT4 | 426.74 |
| 2335 | FN1 | 823.9037 | 6767 | ST13 | 425.5747 |
| 29110 | TBK1 | 672.1386 | 873 | CBR1 | 424.257 |
| 8826 | IQGAP1 | 635.9629 | 667 | DST | 414.0964 |
| 1759 | DNM1 | 628.7167 | 9495 | AKAP5 | 412.7567 |
| 5573 | PRKAR1A | 622.5864 | 8601 | RGS20 | 408.6292 |
| 10397 | NDRG1 | 590.672 | 2896 | GRN | 401.0426 |
| 2923 | PDIA3 | 559.3358 | 4323 | MMP14 | 400.1593 |
| 5339 | PLEC | 541.5909 | 55697 | VAC14 | 394.4482 |
| 6303 | SAT1 | 515.6939 | 5228 | PGF | 393.3167 |
| 2321 | FLT1 | 508.3947 | 8553 | BHLHE40 | 392.8039 |
| 2317 | FLNB | 505.1671 | 604 | BCL6 | 385.7928 |
| 4221 | MEN1 | 499.019 | 5054 | SERPINE1 | 382.5547 |
| 7067 | THRA | 487.6884 | 4597 | MVD | 380.0219 |
| 2147 | F2 | 482.7173 | 5569 | PKIA | 377.6185 |
| 3959 | LGALS3BP | 475.7565 | 140885 | SIRPA | 375.9565 |
| 10603 | SH2B2 | 473.9241 | 3624 | INHBA | 367.5795 |
| 7422 | VEGFA | 465.5212 | 10272 | FSTL3 | 367.2976 |
| 8815 | BANF1 | 461.918 | 4846 | NOS3 | 361.5554 |
| 23650 | TRIM29 | 448.6764 | 3553 | IL1B | 357.2569 |
| 7124 | TNF | 441.7007 | 3880 | KRT19 | 355.4412 |
| 904 | CCNT1 | 440.8304 | 9444 | QKI | 354.3211 |
| 3556 | IL1RAP | 430.4483 | 8773 | SNAP23 | 348.0265 |
| 9131 | AIFM1 | 429.2521 | 27250 | PDCD4 | 345.9363 |
| 6449 | SGTA | 429.1609 | 5291 | PIK3CB | 345.5114 |

The table shown the first 50 genes selected according to the score values and arranged in descendent order. We also provide the gene symbol the associated Entrez Gene identifier.

Notes: 1) Entrez gene ID.

statistical p-value, even when some important and significant inferences, can be made.

A similar situation occurred with the pathway enrichment analysis (Table 3). Even when the KEGG database is the most representative of our gene space and high coherence was noticed with the physiopathology of PE, the results only cover around 50% of the initial 2400 genes. A similar procedure was also performed with the Reactome and BioCarta databases with a less covering (37% and 27% respectively) and showing a high coherence with Table 3 results. These databases reveal other significant pathways like NGF, PDGF, BMP, EPO and EGFR signaling as well as apoptosis and hemostasis pathways (data not shown). Some cancerous pathways (i.e. prostatic, pancreatic and lung) were also found statistically significant in KEGG but were excluded from Table 3, in order to simplify, because many of them have

similar reactions with the general cancer pathway, already presented in the Table 3.

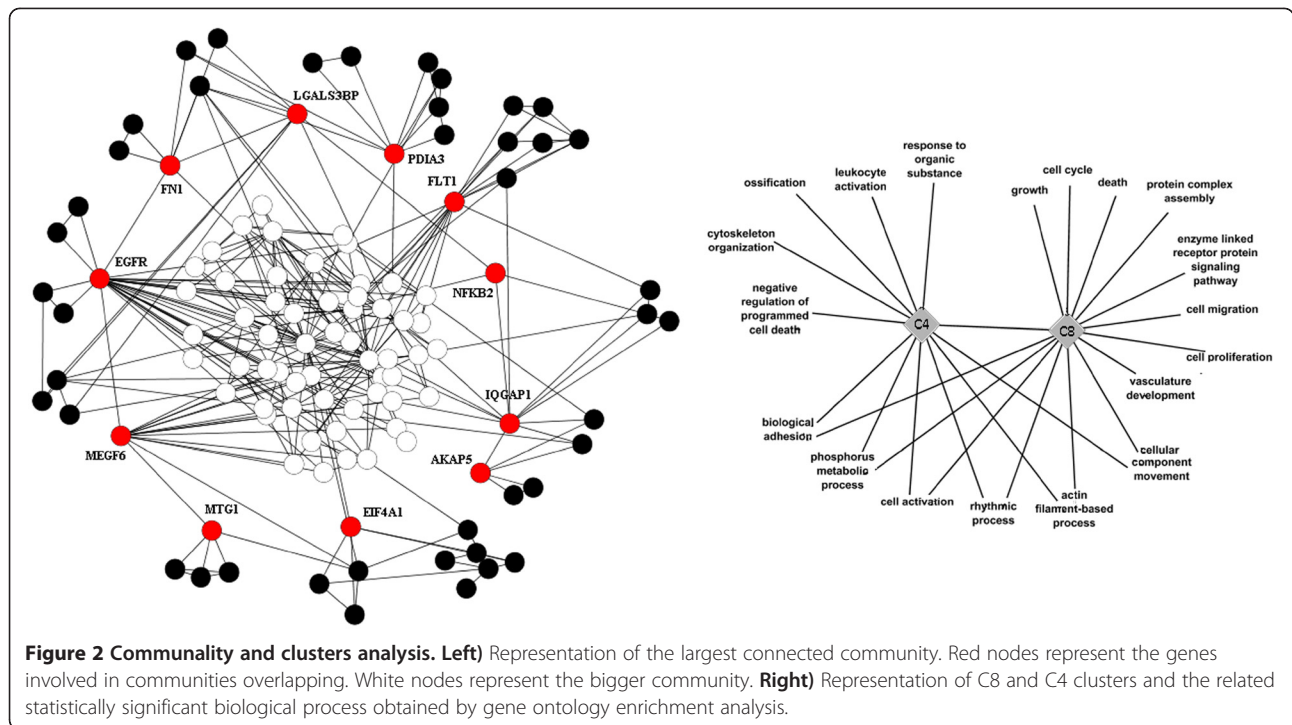
In order to simplify and enhance the understanding of the involved pathways and their relationship with the selected hubs, a fusion between both was made (Figure 3). However, it is important to exalt that from the 50 hubs previously selected; only 22 present some significant pathway association with Table 2. The genes: NDRG1, LGALS3BP, BANF1, SGTA, TRIM29, RGS20, PLEC, GRN, ST13, AKAP5, FSTL3, DST, PKIA, QKI, MLF2 and KRT19, for example, were not found in the KEGG database.

Discussion

PPI analysis

The top 50 genes selected were manually analyzed with the scientific literature in order to validate its connection with PE or even changes during pregnancy and we corroborate that several of them like FLT1, FLT4, VEGFA, PGF, TNF, FN1, F2 and NOS3 can be related to the main lines of research in preeclampsia pathogenesis hypothesis and specially angiogenesis [11-16]. Moreover, PGF, INHBA and related inhibin as well as activin proteins have been considered in several predictive model of PE [17,18]. The presence of those genes in our selection could validate the method applied and increase our confidence with respect to those genes that are poorly explored or unknown in their association with preeclampsia. In the latter group, we have for example the genes LYN, PDIA3, NDRG1 and TBK1. The PDIA3 genes encode an endoplasmic protein highly related with folding processes and in its relation with pregnancy only one article was found [19] revealing that PDIA3 gene intervene in trophoblast invasion via interleukin (IL) 11. Similarly, TBK1 could also intervene during the first moments of pregnancy to secure the implantation in relation with the nuclear factor kappa B [20]. However, there are not articles of PDIA3, TBK1 or LYN related with preeclampsia. Moreover, only one article was found describing an increased expression of NDRG1 during preeclampsia [21]. A similar problem is found with other genes such as IQGAP1, DNM1, SAT1, MEN1 and SH2B2 that also have been little explored during pregnancy.

Cluster analysis indicates that mainly C8 but also C4 probably embrace the most significant genes, at least related with centrality. All genes highest scored are part of these clusters as well as the large community graph and as we can notice the genes that lead to a community superposition are also highly scored in Table 1 (Figure 2). On the other hand there are other genes like ENG, VEGFB and INHA that are well known related to preeclampsia and are also part of the C8 cluster [11-18]. It is important to notice that there are other genes not



shown in Table 1 because were not present in the initial gene set. In this group, EGFR and GRB2, are both with highest scores and there are both well related with preeclampsia [22,23]. Thus, even when the work presented focused on the analysis of the initial group, it is possible that relevant genes were identified by PPI

Table 2 The diseases enrichment analysis

| Disease classes | Disease | p-value |
|------------------------------------|----------------------------|---------------------|
| Cancer (p-value = 1.24E-11) | breast cancer | 9.76E-07 |
| | ovarian cancer | 4.30E-04 |
| | lupus erythematosus | 4.4E-02 |
| Cardiovascular (p-value = 4.28E-9) | myocardial infarction | 8.7E-09 |
| | coronary artery disease | 7.38E-06 |
| | Stroke | 6.16E-05 |
| | heart disease, ischemic | 9.7E-03 |
| | blood pressure, arterial | 8.2E-03 |
| | inflammatory bowel disease | 2.9E-02 |
| | Aging (p-value = 5.23E-05) | Alzheimer's Disease |
| Longevity | | 2.17E-04 |
| atherosclerosis | | 3.7E-02 |
| Arthritis | | 3.60E-02 |
| Reproduction (p-value = 7.9E-03) | preeclampsia | 1.29E-05 |
| | pregnancy loss, recurrent | 6.08E-04 |

The table provides the p-value obtained from disease enrichment analysis according to the Disease Class and also the sub-classes diseases following the GAD database structure. The results showed are partials as discussed in the respective section.

topology but not included in the initial subset. Moreover, the GO enrichment analysis clearly reveals that C8 and C4 clusters are related with angiogenesis, apoptosis and biological adhesion, which are also biological processes obtained with the pathways enrichment analysis. Angiogenesis and related processes (i.e. vascular growing, cell differentiation and hypoxia) are considered as central processes related to preeclampsia [24] and therefore it could support the reliability of the procedure and also the necessity to increase the study of the C8 and C4 genes.

Future combination of centrality indexes and specific clusters selection together with machine learning procedures or genetic algorithm optimization based on groups differentiability or external prediction subset, could reduce even more the final gene space and we are currently working in this direction.

Diseases and pathway analysis

We manually evaluate through scientific literature the connection of PE with the identified metabolic pathways and diseases. The relationship between preeclampsia and inflammatory, immunologic, angiogenic and hemodynamic responses are very well documented [22-28] and therefore are expected not only in the gene space but also in the metabolic pathways analysis. The metabolic pathway analysis (Table 3) indicates a strong significance of the cancer pathway that is consistent with the disease analysis (Table 2) and also with the VEGF signaling, that is present in several related processes like cytokine-cytokine receptor, angiogenesis and also cancer pathway.

Table 3 The KEGG pathway enrichment analysis

| Pathway description | p-value | Pathway description | p-value |
|---|----------|---|----------|
| Pathways in cancer | 1.10E-42 | Fc gamma R-mediated phagocytosis | 2.90E-07 |
| Focal adhesion | 3.50E-29 | RIG-I-like receptor signaling pathway | 5.30E-07 |
| Apoptosis | 2.10E-17 | Chemokine signaling pathway | 5.90E-07 |
| Neurotrophin signaling pathway | 3.40E-17 | Leukocyte transendothelial migration | 6.20E-07 |
| TGF-beta signaling pathway | 5.10E-16 | VEGF signaling pathway | 8.60E-07 |
| T cell receptor signaling pathway | 8.90E-15 | NOD-like receptor signaling pathway | 1.20E-06 |
| ErbB signaling pathway | 1.80E-14 | mTOR signaling pathway | 7.00E-06 |
| B cell receptor signaling pathway | 8.70E-14 | Complement and coagulation cascades | 7.90E-06 |
| Adherens junction | 2.00E-12 | Tight junction | 1.10E-05 |
| Fc epsilon RI signaling pathway | 3.60E-12 | RNA polymerase | 1.60E-05 |
| ECM-receptor interaction | 6.80E-12 | Ubiquitin mediated proteolysis | 2.20E-05 |
| Insulin signaling pathway | 3.00E-10 | GnRH signaling pathway | 1.10E-04 |
| Regulation of actin cytoskeleton | 3.10E-09 | Hematopoietic cell lineage | 1.20E-04 |
| Toll-like receptor signaling pathway | 1.01E-08 | Cytokine-cytokine receptor interaction | 2.00E-04 |
| Wnt signaling pathway | 1.10E-08 | Hedgehog signaling pathway | 3.70E-04 |
| Cell cycle | 1.50E-08 | Natural killer cell mediated cytotoxicity | 3.80E-04 |
| MAPK signaling pathway | 1.50E-08 | Cytosolic DNA-sensing pathway | 6.70E-04 |
| Adipocytokine signaling pathway | 7.50E-08 | Notch signaling pathway | 2.60E-03 |
| Progesterone-mediated oocyte maturation | 2.50E-07 | Renin-angiotensin system | 1.80E-02 |

The table provides the p-value obtained from pathway enrichment analysis in the KEGG database. The pathways names correspond with the KEGG nomenclature. The results showed are partials as discussed in the respective section.

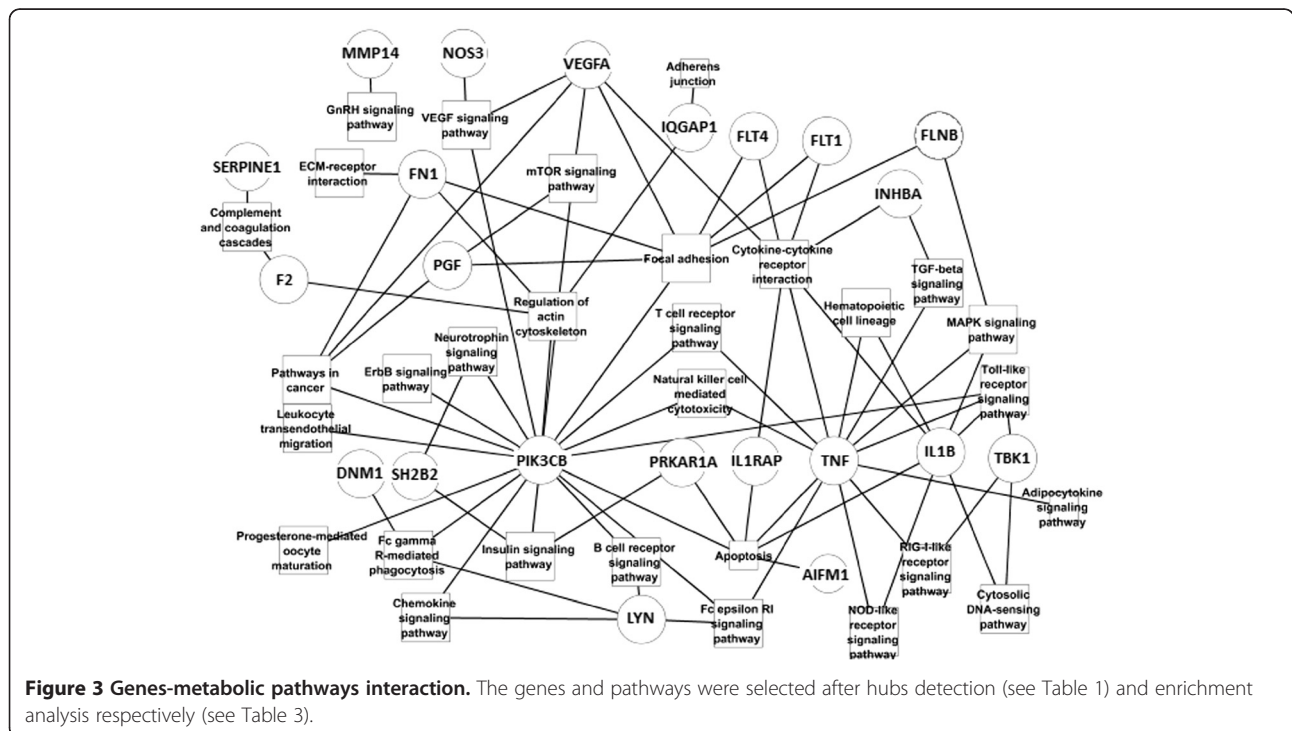


Figure 3 Genes-metabolic pathways interaction. The genes and pathways were selected after hubs detection (see Table 1) and enrichment analysis respectively (see Table 3).

With the procedure carried out in the present study a simplification of the metabolic pathways was possible, however, more needs to be done in this area in order to better integrate not only the hubs genes but also the comprehensive data created by the PPI interaction.

Considering the associated metabolic pathways we can notice that several signaling pathways are statistically significant and some of them are connected with PIK3CB, LYN and linked in several cases with TNF (Figure 3) that is a central gene affecting several processes and also widely studied in its relation with preeclampsia [28]. Similarly, SH2B2 that encodes an adapter protein of several tyrosine kinase receptors is also connected with metabolic pathways indirectly related with apoptosis. Contrasting PDIA3 and NDRG1 are not present in the Figure 3 connected with any metabolic pathway, but the recent relation found between PDIA3 and IL11 could open a relationship with cytokine-cytokine receptors [19], specifically through the hematopoietins receptor family. We can also notice (Figure 3 and Table 3) that in the pathways associated with central proteins (the hubs), those highly connected are the cytokine-cytokine receptors, focal adhesion and apoptosis pathways and they contain almost the complete space of genes mainly studied in preeclampsia.

The elevated number of signaling pathways that we found statistically significant, together with the hubs distribution detected in Table 1, seems to point out the idea of a signaling related disease similar to cancer, however, the apoptosis and angiogenic mechanism had been related previously with PE and are also highly represented in our study.

On the other hand, the relationship between PE and cardiovascular diseases is well known. Women with previous history of PE or even pregnancy hypertension present an increased risk of future development of cardiovascular disease or hypertension [29-31]. This is clearly expressed in our results (Table 2). Moreover, our results also suggest a significant relationship between cancer and PE (at this point it is not possible to say if this correlation is positive or negative) indicated by diseases enrichment analysis and also by the metabolic pathways. However, the experimental and epidemiological evidence that could support the influence of PE in future cancer development is for now inconclusive, even when could be reasonably expected by the wide kind of genes that are actually shared between both diseases (i.e. angiogenesis related genes). Several articles report a reduced risk of future cancer development in women with previous history of PE, but others could not find any statistical relationship [32-35].

There are neither clinical signs in common between PE and Alzheimer and nor epidemiological studies. However, the connection between both diseases has been questioned

before (at least in late-onset Alzheimer's disease) [36] and was significant in the present study. Therefore, although it is not possible to validate our findings experimentally with the current information, it actually opens new possibilities for future works. A similar problem concerns to other ageing related factors like atherosclerosis, arthritis or longevity (that for obvious reasons will be difficult to explore in PE during long-term studies).

The present study points out several advantages of the bioinformatics analysis but some limitations were also found. The detection of genes that are very well known to be involved in PE thought the applied methodology as well as the consistence of the results across metabolic information could support the novel candidates found, however, it is necessary to reduce even more the genes space applying other methodologies as well as to design new experimental experiences. On the other hand, the limitation of the human protein interaction information suggests that also orthologous genes should be needed in order to increase the PPI covering of the initial dataset and to increase the capabilities of the metabolic pathways and disease enrichment analysis.

Conclusions

In the present study we applied several bioinformatics tools in order to create a comprehensive initial database of genes statistically related to preeclampsia and a further expansion through the construction of related protein-protein interaction network. Using several centrality indexes for hubs detection, some well know preeclampsia related genes like FLT1, TNF, VEGFA and PGF were detected as well as other genes with high scores, like PDIA3, NDRG1, TBK1, LYN, IQGAP1, DNMT1, SAT1, MEN1 and SH2B2 that have been poorly explored or unknown in the current state of the art of preeclampsia pathophysiology.

Through disease enrichment analysis we corroborated the well know connection between PE and cardiovascular disease, but we also found a possible link of PE with aging and cancer. Moreover we also found the cancer pathway, focal adhesion, ECM-receptor interaction, apoptosis and cytokine-cytokine receptors interactions metabolic pathways highly represented by the PPI network that are in agreement with some of the preeclampsia-diseases relationship found, as well as with the central topics of the current preeclampsia investigations.

Methods

Dataset construction

Experimental proteomic/genomic data comparing normal (N) and preeclamptic (PE) pregnancies was obtained analysing the Gene Expression Omnibus (GEO) database [37]. The datasets considered are represented in Table 4.

Table 4 GEO data used for dataset construction

| GEO | Sample | Method | Collection (Weeks) | Tissue | Ref. |
|-----------------------|---------------|------------|--------------------|----------|------|
| GSE6573 | 10(PE), 10(N) | Affymetrix | At delivery | Placenta | [25] |
| GSE4707 ¹ | 10(PE), 4(N) | Agilent | See notes | Placenta | [6] |
| GSE25906 | 23(PE), 37(N) | Illumina | 27-38 | Placenta | [26] |
| GSE14722 ² | 12(PE), 11(N) | Affymetrix | See notes | Placenta | [27] |
| GSE10588 | 17(PE), 26(N) | ABI Human | At delivery | Placenta | [38] |

The Table shown the GEO data sources, references and complementary information used in the study.

Notes: 1) The 10 samples were collected for placenta biopsy during delivery. Five of the preeclamptic samples were collected before 31 weeks of gestation and the rest after this time. 2) The 23 samples were collected for placenta biopsy during delivery, however, because in GEO appear two platform types, both were analysed.

Each experiment was analyzed independently in order to reduce the number of genes. In our case we considered an adjusted p-value ≤ 0.05 and a fold expression ≥ 2 as discussed elsewhere [6,7,25-27,38,39]. Initially the p-value was obtained by a bootstrapping procedure with 1000 or 10000 iterations (depending on size of the sample) obtaining 645 statistically significant modulated genes, however, applying the false discovery rate (FDR) correction by the Benajmini-Hochberg method [40], this sample was reduced to 330 genes.

In addition, several text mining exploration tools were used to complement the GEO results. There are several tools to perform a text data mining analysis but several of them require extra information (i.e. chromosome region) instead phenotype or diseases notation (i.e. diseases name or related keywords). In our case we choose those methods that do not require previous genetic knowledge of the disease [8]. Moreover, the text mining procedures usually could provide several false positive associations and therefore those tools which also combine text-mining with other data sources in the analysis are preferred [8,41]. Considering these aspects, we used the following tools: PolySearch [42], Candid [43] and Phenopred [44]. Candid and PhenoPred use several heterogeneous data sources to overcome bias while in PolySearch analyse was restricted to PubMed publications. Obviously many other algorithms could be used in alternative. In order to reduce the risk of including biased relationships, the top 10–20 genes/proteins with the highest scores were selected and individually analyzed considering the preeclampsia related scientific publications. Some of the top genes were also present in the previous dataset (GEO), therefore, the final dataset contained 347 genes.

Protein-protein interaction network (PPI)

The proteins associated with the previous 347 genes were identified and cross-referenced with the IRefIndex

(v1.16) [45] and a signaling curated databases [46] that were used to create the protein-protein interaction (PPI) network. The IRefIndex database provide an index of protein interactions available in several databases like: BIND, BioGRID, DIP, HPRD that simplify the task consuming process of inter-database mapping and lead to a comprehensive covering of the available known protein interactions space. On the other hand, this PPI database is easily integrated in Cytoscape. Additionally, many diseases are related with signaling pathways modifications and therefore the inclusion of this interaction database considerable improve the PPI space. The interaction search was restricted to *Homo Sapiens* and includes all kind of experimental procedure as well as some predictive interactions (mainly from the OPHID database). The curation of the final database was performed both, manually and using home-made software to remove duplicate interactions and unify isoforms notation with unique genes. We obtained our final PPI network with 3279 interactions and 2400 nodes.

Some of the proteins present in our initial dataset had not any known experimental interaction (at least in humans) and therefore the 2400 nodes cover only 234 (67.45%) genes of the initial set (347). The network visualization and network topology indexes, calculated in the hubs detection process, were carried out using Cytoscape 2.8.2 and CytoHubba [47,48].

Several methodologies are available for hubs and essential genes identification, and all of them with the respective advantage and limitations [47,49-54]. Some strategies are the use of genetic algorithm or machine learning procedures [49,50], however, the centrality approaches are by far the most applied procedures even by simplicity and because several studies had being pointed out its applicability [47,51,52]. Therefore, several centrality indexes were evaluated: Betweenness, bottleneck, density of maximum neighborhood component (DMNC), node degree, edge percolation component (EPC), eccentricity, maximal clique centrality (MCC), maximum neighborhood component (MNC), radiality and stress [47]. On the other hand, to obtain a scoring index we created the measurement (*Score I*) as follow:

$$ScoreI = \sum_{i=1}^{Nc} \left[\left(1 - \frac{\max(Ic_i) - Ic_i}{\max(Ic_i) - \min(Ic_i)} \right) \cdot 100 \right] \quad (1)$$

Where Ic_i is the values of centrality indexes and $i = 1 \dots Nc$, and is the number of calculated centrality indexes ($Nc = 10$). As we can note, *score I* is the sum of all the indexes percent value after individual normalization and therefore is restricted to a maximal value of $100 \times Nc$ which simplify even better the top genes selection. With the normalized centrality indexes we also performed a model-based clustering analysis using R-package [53] in

order to study hubs distribution with respect to centrality ranks. We also performed a communality (or cliques) network analysis by clique percolation method using CFinder [54]. The communality analysis provides a better topology description of the network including the location of highly connected sub-graphs (cliques) and/or overlapping modules that usually correspond with relevant biological information.

Pathway and diseases enrichment analysis

The pathways and diseases enrichment analysis were performed through the DAVID bioinformatics resource 6.7 [55], exploring the well know databases: KEGG, BioCarta and Reactome (pathways related) as well as OMIN and Genetic Association Database (GAD) (diseases related analysis). This online resource (DAVID) integrate, in a faster computational analysis, a wide range of enrichment analysis thought different databases providing also a substantial statistical description. The analysis was carried out considering the complete gene space of the PPI network. We also used DAVID in order to perform a gene ontology enrichment analysis in the obtained clusters.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ET implemented the methods and worked out in the database processing. ET, JB and IR designed the study as well as worked out in the biological background for the discussions. All authors contributed to the writing of the manuscript and read and approved the manuscript.

Acknowledgements

This study was supported by "Fundação para a Ciência e a Tecnologia" (FCT), grant: SFRH/BPD/72391/2010.

Author details

¹Department of Biological Sciences, Biochemistry, Faculty of Pharmacy, University of Porto, Portugal/Institute for Molecular and Cell Biology (IMC), Porto, Portugal. ²Department of Obstetrics and Gynecology, Faculty of Medicine, University of Porto, Department of Obstetrics and Gynecology, Sao Joao Hospital./INEB – Institute of Biomedical Engineering, University of Porto, Porto, Portugal.

Received: 5 January 2012 Accepted: 23 July 2012

Published: 8 August 2012

References

1. Leeman L, Fontaine P: Hypertensive disorders of pregnancy. *Am Fam Physician* 2008, **78**(1):93–100.
2. National Collaborating Centre for Women's and Children's Health: *Hypertension in pregnancy. The management of hypertensive disorders during pregnancy*. London (UK): National Institute for Health and Clinical Excellence (NICE); 2010:46. Clinical guideline; no. 107.
3. Sheppard SJ, Khalil RA: Risk factors and mediators of the vascular dysfunction associated with hypertension in pregnancy. *Cardiovasc Hematol Disord Drug Targets* 2010, **10**(1):33–52.
4. Borna S, Neamatipoor E, Radman N: Risk of coronary artery disease in women with history of pregnancies complicated by Preeclampsia and LBW. *J Matern Fetal Neonatal Med* 2011, **19**. Epub ahead of print.
5. Carty DM, Delles C, Dominiczak AF: Preeclampsia and future maternal health. *J Hypertens* 2010, **28**(7):1349–1355.
6. Nishizawa H, Pryor-Koishi K, Kato T, Kowa H, Kurahashi H, Udagawa Y: Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia. *Placenta* 2007, **28**(5–6):487–497.
7. Founds SA, Conley YP, Lyons-Weiler JF, Jeyabalan A, Hogge WA, Conrad KP: Altered global gene expression in first trimester placentas of women destined to develop Preeclampsia. *Placenta* 2009, **30**(1):15–24.
8. Tranchevent LC, Capdevila FB, Nitsch D, Moor BD, Causmaecker PD, Moreau Y: A guide to web tools to prioritize candidate genes. *Brief Bioinform* 2010, **12**(1):22–32.
9. Özgür A, Vu T, Erkan G, Radev DR: Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 2008, **24**:i277–i285.
10. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE: A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005, **122**(6):957–968.
11. Ward K: Searching for genetic factors underlying pre-eclampsia: recent progress and persistent challenges. *Minerva Ginecol* 2008, **60**(5):399–419.
12. Nejatizadeh A, Stobdan T, Malhotra N, Pasha MA: The genetic aspects of pre-eclampsia: achievements and limitations. *Biochem Genet* 2008, **46**(7–8):451–479.
13. Laresgoiti-Servitje E, Gómez-López N, Olson DM: An immunological insight into the origins of pre-eclampsia. *Hum Reprod Update* 2010, **16**(5):510–524.
14. Winn VD, Gormley M, Fisher SJ: The impact of Preeclampsia on gene expression at the maternal-fetal interface. *Pregnancy Hypertens* 2011, **1**(1):100–108. 1.
15. Ishida J, Matsuoka T, Saito-Fujita T, Inaba S, Kunita S, Sugiyama F, Yagami K, Fukamizu A: Pregnancy-associated homeostasis and dysregulation: lessons from genetically modified animal models. *J Biochem* 2011, **150**(1):5–14.
16. Rajakumar A, Chu T, Handley DE, Bunce KD, Burke B, Hubel CA, Jeyabalan A, Peters DG: Maternal gene expression profiling during pregnancy and preeclampsia in human peripheral blood mononuclear cells. *Placenta* 2011, **32**(1):70–78.
17. Kuc S, Wortelboer EJ, van Rijn BB, Franx A, Visser GH, Schielen PC: Evaluation of 7 serum biomarkers and uterine artery Doppler ultrasound for first-trimester prediction of preeclampsia: a systematic review. *Obstet Gynecol Surv* 2011, **66**(4):225–239.
18. Spencer K, Yu CK, Savvidou M, Papageorgiou AT, Nicolaides KH: Prediction of pre-eclampsia by uterine artery Doppler ultrasonography and maternal serum pregnancy-associated plasma protein-A, free beta-human chorionic gonadotropin, activin A and inhibin A at 22 + 0 to 24 + 6 weeks' gestation. *Ultrasound Obstet Gynecol* 2006, **27**(6):658–663.
19. Sonderegger S, Yap J, Menkhorst E, Weston G, Stanton PG, Dimitriadis E: Interleukin (IL)11 mediates protein secretion and modification in human extravillous trophoblasts. *Hum Reprod* 2011, **26**(10):2841–2849.
20. King AE, Critchley HO, Kelly RW: The NF-kappaB pathway in human endometrium and first trimester decidua. *Mol Hum Reprod* 2001, **7**(2):175–183.
21. Choi SJ, Oh SY, Kim JH, Sadovsky Y, Roh CR: Increased expression of N-myc downstream-regulated gene 1 (NDRG1) in placentas from pregnancies complicated by intrauterine growth restriction or preeclampsia. *Am J Obstet Gynecol* 2007, **196**(1):45e1–7.
22. Anteby EY, Ayesb S, Shochina M, Hamani Y, Schneider T, Al-Shareef W, Hochberg A, Ariel I: Growth factor receptor-protein bound 2 (GRB2) upregulation in the placenta in preeclampsia implies a possible role for ras signalling. *Eur J Obstet Gynecol Reprod Biol* 2005, **118**(2):174–181. 1.
23. Faxén M, Nasiell J, Blanck A, Nisell H, Lunell NO: Altered mRNA expression pattern of placental epidermal growth factor receptor (EGFR) in pregnancies complicated by preeclampsia and/or intrauterine growth retardation. *Am J Perinatol* 1998, **15**(1):9–13.
24. Verlohren S, Stepan H, Dechend R: Angiogenic growth factors in the diagnosis and prediction of pre-eclampsia. *Clin Sci (Lond)* 2012, **122**(2):43–52.
25. Herse F, Dechend R, Harsem NK, Wallukat G, Janke J, Qadri F, Hering L, Muller DN, Luft FC, Staff AC: Dysregulation of the circulating and tissue-based renin-angiotensin system in preeclampsia. *Hypertension* 2007, **49**(3):604–611.

26. Tsai S, Hardison NE, James AH, Motsinger-Reif AA, Bischoff SR, Thames BH, Piedrahita JA: **Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals dysregulation of sialic acid acetyltransferase and immune signalling pathways.** *Placenta* 2011, **32**(2):175–182.
27. Winn VD, Gormley M, Paquet AC, Kjaer-Sorensen K, Kramer A, Rumer KK, Haimov-Kochman R, Yeh RF, Overgaard MT, Varki A, Oxvig C, Fisher SJ: **Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2.** *Endocrinology* 2009, **150**(1):452–462.
28. Xie C, Yao MZ, Liu JB, Xiong LK: **A meta-analysis of tumor necrosis factor- α , interleukin-6, and interleukin-10 in preeclampsia.** *Cytokine* 2011, **6**(3):550–559.
29. Portelinha A, Cerdeira AS, Belo L, Braga J, Tejera E, Pinto A, Pinto F, Areias MJ, Patrício B, Rebelo I: **Haemostatic factors in women with history of preeclampsia.** *Thromb Res* 2009, **124**(1):52–56.
30. Williams D: **Long-term complications of preeclampsia.** *Semin Nephrol* 2011, **31**(1):111–122.
31. Roberts JM, Hubel CA: **Pregnancy: a screening test for later life cardiovascular disease.** *Womens Health Issues* 2010, **20**–5:304–307.
32. Ma H, Henderson KD, Sullivan-Halley J, Duan L, Marshall SF, Ursin G, Horn-Ross PL, Largent J, Deapen DM, Lacey JV Jr, Bernstein L: **Pregnancy-related factors and the risk of breast carcinoma *in situ* and invasive breast cancer among postmenopausal women in the California Teachers Study cohort.** *Breast Cancer Res* 2010, **12**(3):R35.
33. Calderon-Margalit R, Friedlander Y, Yanetz R, Deutsch L, Perrin MC, Kleinhaus K, Tiram E, Harlap S, Paltiel O: **Preeclampsia and subsequent risk of cancer: update from the Jerusalem Perinatal Study.** *Am J Obstet Gynecol* 2009, **200**(1):63.e1–5.
34. Vatten LJ, Romundstad PR, Jennum PA, Eskild A: **Angiogenic balance in pregnancy and subsequent breast cancer risk and survival: a population study.** *Cancer Epidemiol Biomarkers Prev* 2010, **20**:9:18.
35. Nechuta S, Paneth N, Velie EM: **Pregnancy characteristics and maternal breast cancer risk: a review of the epidemiologic literature.** *Cancer Causes Control* 2010, **21**(7):967–989.
36. van Dijk M, van Bezu J, Poutsma A, Veerhuis R, Rozemuller AJ, Scheper W, Blankenstein MA, Oudejans CB: **The pre-eclampsia gene STOX1 controls a conserved pathway in placenta and brain upregulated in late-onset Alzheimer's disease.** *J Alzheimers Dis* 2010, **19**(2):673–679.
37. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muetter RN, Holko M, Ayhanbule O, Yefanov A, Soboleva A: **NCBI GEO: archive for functional genomics data sets-10 years on.** *Nucleic Acids Res* 2011, **39**(Database issue):D1005–D1010.
38. Sitras V, Paulssen RH, Grønnaas H, Leirvik J, Hanssen TA, Vårtun A, Acharya G: **Differential placental gene expression in severe preeclampsia.** *Placenta* 2009, **30**(5):424–433.
39. Guda P, Chittur SV, Guda C: **Comparative analysis of protein-protein interactions in cancer-associated genes.** *Genomics Proteomics Bioinformatics* 2009, **7**(1–2):25–36.
40. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Statist Soc Ser B* 1995, **57**:289–300.
41. Mahdavi MA, Lin YH: **False positive reduction in protein-protein interaction predictions using gene ontology annotations.** *BMC Bioinforma* 2007, **23**:8–262.
42. Cheng D, Knox C, Young N, Stothard P, Damaraju S, Wishart DS: **PolySearch: a webbased text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites.** *Nucleic Acids Res* 2008, **36**:W399–W405.
43. Hutz JE, Kraja AT, McLeod HL, Province MA: **CANDID: a flexible method for prioritizing candidate genes for complex human traits.** *Genet Epidemiol* 2008, **32**(8):779–790.
44. Radivojac P, Peng K, Clark WT, Peters BJ, Mohan A, Boyle SM, Mooney SD: **An integrated approach to inferring gene-disease associations in humans.** *Proteins* 2008, **72**(3):1030–1037. 15.
45. Turner B, Razick S, Turinsky AL, Vlasblom J, Crowley EK, Cho E, Morrison K, Donaldson IM, Wodak SJ: **iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence.** Oxford: Database; 2010, 12.
46. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, Purisima EO, Wang E: **A map of human cancer signaling.** *Mol Syst Biol* 2007, **3**:152.
47. Chung-Yen L, Chia-Hao C, Hsin-Hung W, Shu-Hwa C, Chin-Wen H, Ming-Tat K: **Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology.** *Nucleic Acids Res* 2008, **36**:w438–w443.
48. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2011, **27**(3):431–432.
49. Kong W, Mou X, Hu X: **Exploring matrix factorization techniques for significant genes identification of Alzheimer's disease microarray gene expression data.** *BMC Bioinforma* 2011, **12**(Suppl 5):S7.
50. Mishra H, Singh N, Misra K, Lahiri T: **An ANN-GA model based promoter prediction in *Arabidopsis thaliana* using tilling microarray data.** *Bioinformatics* 2011, **6**(6):240–243.
51. Rio G, Koschützki D, Coello G: **How to identify essential genes from molecular networks?** *BMC Syst Biol* 2009, **3**:102.
52. Pang K, Sheng H, Mab X: **Understanding gene essentiality by finely characterizing hubs in the yeast protein interaction network.** *Biochem Biophys Res Commun* 2010, **401**:112–116.
53. Fraley C, Raftery AE: **MCLUST: software for model-based cluster analysis.** *J Classif* 1999, **16**:297–306.
54. Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T: **CFinder: locating cliques and overlapping modules in biological networks.** *Bioinformatics* 2006, **22**:1021–1023.
55. Huang DW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1–13.

doi:10.1186/1752-0509-6-97

Cite this article as: Tejera et al.: Preeclampsia: a bioinformatics approach through protein-protein interaction networks analysis. *BMC Systems Biology* 2012 **6**:97.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

