

RESEARCH ARTICLE

Open Access



Consensus strategy in genes prioritization and combined bioinformatics analysis for preeclampsia pathogenesis

Eduardo Tejera^{1*} , Maykel Cruz-Monteagudo^{2,3,6,7}, Germán Burgos¹, María-Eugenia Sánchez¹, Aminaél Sánchez-Rodríguez⁴, Yunierkis Pérez-Castillo⁵, Fernanda Borges⁶, Maria Natália Dias Soeiro Cordeiro⁷, César Paz-y-Miño⁸ and Irene Rebelo^{9,10}

Abstract

Background: Preeclampsia is a multifactorial disease with unknown pathogenesis. Even when recent studies explored this disease using several bioinformatics tools, the main objective was not directed to pathogenesis. Additionally, consensus prioritization was proved to be highly efficient in the recognition of genes-disease association. However, not information is available about the consensus ability to early recognize genes directly involved in pathogenesis. Therefore our aim in this study is to apply several theoretical approaches to explore preeclampsia; specifically those genes directly involved in the pathogenesis.

Methods: We firstly evaluated the consensus between 12 prioritization strategies to early recognize pathogenic genes related to preeclampsia. A communality analysis in the protein-protein interaction network of previously selected genes was done including further enrichment analysis. The enrichment analysis includes metabolic pathways as well as gene ontology. Microarray data was also collected and used in order to confirm our results or as a strategy to weight the previously enriched pathways.

Results: The consensus prioritized gene list was rationally filtered to 476 genes using several criteria. The communality analysis showed an enrichment of communities connected with VEGF-signaling pathway. This pathway is also enriched considering the microarray data. Our result point to VEGF, FLT1 and KDR as relevant pathogenic genes, as well as those connected with NO metabolism.

Conclusion: Our results revealed that consensus strategy improve the detection and initial enrichment of pathogenic genes, at least in preeclampsia condition. Moreover the combination of the first percent of the prioritized genes with protein-protein interaction network followed by communality analysis reduces the gene space. This approach actually identifies well known genes related with pathogenesis. However, genes like HSP90, PAK2, CD247 and others included in the first 1% of the prioritized list need to be further explored in preeclampsia pathogenesis through experimental approaches.

Keywords: Consensus analysis, Gene periodization, Preeclampsia, Communality analysis, Microarray analysis, Pathogenesis, Early recognition

* Correspondence: edutp00@gmail.com

¹Facultad de Medicina, Universidad de Las Américas, Av. de los Granados E12-41y Colimes esq, EC170125 Quito, Ecuador

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



Background

The study of preeclampsia (PE) from a bioinformatics approach will be affected by several aspects that will inevitably affect the interpretation and will establish an implicit frame to our analysis. The PE is a multifactorial disease that probably involves several genes and environmental factors. However, the main theory behind PE is that the disorder results from placenta ischemia, with further releases of several factors into the maternal circulation [1, 2]. The ischemia origin is supported mainly by a failure in the transformation of the spiral artery caused by a trophoblastic invasion abnormality [1–3]. Therefore, placenta (at this level) is the central organ for pathogenesis. From this point forward, the possible scenarios could be more complex. Nevertheless, the endothelial dysfunction seems to be a primary factor leading to the remaining problems and clinical manifestations. The roll of placenta is clearly reflected in the application of “omic” tools, specifically microarrays studies. The simple inspection of microarray data [4, 5] (through GEO and ArrayExpress databases) reveals that majority were obtained in placenta samples by a case/control design.

Even when arrays technologies could be valuable to provide a wide gene-disease association, a problem arises from the experimental design (case/control). With this type of experimental design will be hard to differentiate pathogenic from non-pathogenic genes. Its means that if we obtain a very significant up or down-regulated gene, we can't be sure that it is involved in pathogenesis. Moreover we can't confirm that these up or down-regulated genes can be used as a risk evaluator or predictive measure without further experimental analysis in a longitudinal design. Even with all previous considerations, microarrays information is used for bioinformatics analysis and gene prioritization suggesting that here are genes that can be probably related with pathogenesis [6–12].

How is the scenery in scientific literature? The ratio between case/control and prospective analysis is biased. It is difficult to proof this statement without a rigorous analysis of the scientific information. However, using pre-eclampsia (MeSH Term) in PubMed database, we obtain 13,173 publications in the last 10 years, but adding the terms “longitudinal studies” or “prospective” the previous search is reduced to 1578 in the same time interval. Even when this approach can be considered as superficial clearly indicate the bias toward case/control studies. Therefore, any prioritization strategy based on text mining or even database exploration, will provide us with a genetic-disease association. However we can't confirm that there genes are primarily related to PE pathogenesis.

There are few studies in PE focused in system biology or other bioinformatics tools [6–14]. Some of them use the microarray information previously described while

others used (combined or not), text mining and protein-protein interaction networks (PPI). All these methods will be affected by the previous discussed issues. Still, a more important problem with bioinformatics tools is actually its diversity. There are several ways in which we can combine the information and not all of them will converge into the same results. For example, in the recent work of Miranda van Uitert et al. [6] on microarray data, their proposed several genes but when compared with other two similar studies we found an overlap of 77% with Vaiman et al. [13] and 44% with Moslehi et al [14]. Between these three studies a total of 556 genes were selected but only 47 are common (~8%) which is a very low overlapping (this microarray information will be further discussed).

Each particular problem could have a better tool to solve it and in terms of prioritization, the consensus strategy had proof to be the most effective way to explore gene-disease association [15, 16]. However, we are not so clear if consensus is also effective for identification of pathogenic genes and it will be our first step in the current work. Consequently we will include several prioritization strategies that will be integrated using a consensus strategy in order to rank the genes in the gene-disease association. The consensus result will be integrated in a common pathway and compared with previous results in microarray meta-analysis in order to clarify the genetic function. The goal of this second step including network analysis and metabolic pathways analysis is to additionally evaluate the capacity of identify pathogenic pathways and its relevance.

Methods

Selection of pathogenic genes for validation

In order to validate the prioritization strategy on pathogenic genes we need to identify specific genes with a high probability of being involved in PE pathogenesis. Through manually literature inspection we considered a gene as pathogenic if:

1. The silencing or induced overexpression of the proposed gene in animal models generate a clinical phenotype like preeclampsia (this group of genes was named as G1)
2. At least one variation (polymorphism) was associated with PE. We only consider the articles that apply meta-analysis methods (this group of genes was named as G2)

The full analysis of the genes in each group can be found in Additional file 1. We found 35 unique genes combining G1 and G2 groups (off course it is not an exhaustive list). The selected genes in each group and its corresponding Entrez Gene ID identifier are:

G1 ($n = 27$): ADA (100), ADORA2A (135), ADORA2B (136), AGTR1 (185), APOH (350), CD73 (4907), CRP (1401), ENG (2022), EDN1 (1606), FLT1 (2321), GADD45A (1647), HADHA (3030), HIF1A (3091), IDO1 (3620), IL10 (3586), IL17A (3605), IL6 (2569), NOS1 (4842), NOS2 (4843), NOS3 (4846), PGF (5228), ROS1 (6098), TACR3 (6870), TGFB1 (7040), TNF (7124), TNFSF14 (8740), VEGFA (7422).

G2 ($n = 13$): F5 (2153), F2 (2147), AGT (183), MTHFR (4524), NOS3 (4846), ACE (1636), SERPINE1 (5054), VEGFA (7422), LEPR (3953), TGFB1 (7040), AGTR1 (185), HLA-G (3135), IL10 (3586).

Prioritization algorithms and consensus strategy

From the prioritization portal [15, 16] we selected the methods according to the following criteria: 1) fully available in web service platform and 2) requiring only the disease name for gene prioritization. Under these conditions we found 12 methods: Biograph [17], Candid [18], Glad4U [19], PolySearch [20], CIPHER [21], Guildify [22], DisgeNet [23], GeneProspector [24], Genie [25], SNPs3D [26], GeneDistiller [27] and MetaRanker [28]. The following methods: CIPHER, Guildify and DisgeNet were not selected from the prioritization portal but from literature, however, fulfilling the same two requirements. These methods have several characteristics that had being fully comprised by other authors previously [15, 29].

Our strategy to combine the different scores obtained in each independent methods is similar to the method used in [30, 31]. Each gene (denoted as i) in the ranked list provided by each method (denoted as j) was normalized ($GeneN_{i,j}$ which means, the normalized score of the gene “ i ” in the method “ j ”) in order to integrate all methods for the consensus approach. For the final score by gene, we considered the average normalized score as well as the number of methods which predict the gene (denoted as n_i) using the formula:

$$Gene_i = \sqrt{\left(\frac{n_i-1}{12-1}\right) \left(\frac{1}{j} \sum_j GeneN_{i,j}\right)} \tag{1}$$

The Eq. 1 correspond with the geometrical mean between the average score of each gene obtained in each method and the normalized score according to the number of methods which predict the association between the gene and the disease. However, this formula will be zero if the gene is only predicted by only one method. Therefore we sort the genes according to the $Gene_i$ values and according to the average $\left(\left[\left(\frac{n_i-1}{12-1}\right) + \left(\frac{1}{j} \sum_j GeneN_{i,j}\right)\right] / 2\right)$. This sorting will produce a ranking that further normalized leading to the final score of each gene ($ConsenScore_i$). If two genes are predicted by only one method and also have the same

normalized scores then will also have the same value of $ConsenScore_i$.

The final list of prioritized genes is actually very long (more than 18,000). We needed a strategy to create a rational cutoff in the number of genes comprising the major pathogenic information with minimal noise. To accomplish this task we used the same pathogenic genes already defined. We defined the following index: $I_i = \frac{TP_i}{FP_i+1}$ $ConsenScore_i$ where, TP and FP are the true and false positive values (up to the ranking value of the gene i) respectively. The maximal value of I_i can be understood as the maximal compromise between the true positive and false positive rate compensated with the ranking index of each gene.

Early recognition analysis in prioritization

Several enrichment metrics have been proposed in the chemoinformatics literature to measure the enrichment ability of a virtual screening protocol [32] and had being recently applied in gene prioritization [33]. In this work and similar to [33], we used some of the most extended metrics to estimate the enrichment ability in order to compare different gene prioritization strategies. The overall enrichment metrics include the area under the accumulation curve ($AUAC$); the area under the ROC curve (ROC); and the enrichment factor (EF) evaluated at the top 1, 5, 10 and 20% of the ranked list. At the same time, the early recognition metrics used were the robust initial enhancement (RIE) and the Boltzmann-enhanced discrimination of ROC ($BEDROC$) evaluated at the top 1, 5, 10 and 20% of the ranked list [32]. The calculation of both classic and early recognition enrichment metrics was conducted by using the perl script *Cresset_VS* [34].

Enrichment analysis

We used David Bioinformatics Resource [35, 36] for gene ontology (GO) and pathways enrichment analysis. The number of GO terms could be very big considering the amount of genes. Therefore we used Revigo [37] in order to simplify the GO terms keeping those with highest specificity. We additionally used RSpider [38], to obtain an integrated pathway combining Reactome and KEGG databases. In these databases the pathways are not the same so any enrichment will produce different pathways that otherwise could be connected or even very similar in the two databases. The use of RSpider will produce not only a statistical analysis of the enrichment but also a network representation integrating the information in both databases. The main goal in RSpider is to connect into non-interrupted sub-network component as many input genes as possible using minimal number of missing genes.

Protein-protein interaction network and analysis

We used String Database [39] to create the protein-protein interactions network with a confidence cutoff of 0.9 and zero node addition. We also used Cytoscape [40] for centrality indexes calculation and network visualization.

Communality (or cliques) network analysis by clique percolation method was applied using CFinder [41]. 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. The selection of the value “k-cliques” will affect the number of community and also the number of genes in each community. We create a rational cutoff by balancing the number of communities and the genes distribution across them. In general higher values of k-cliques imply few communities while lower values lead to many communities. In our network both extremes (too small or to high k-cliques values) result in an unbalanced distribution of the genes across communities. Therefore we create the following index “S” as: $S^k = \frac{mean(N_g^k) - median(N_g^k)}{N_c^k}$ where N_g^k and N_c^k are the number of genes in each community and the number of communities for a defined k-clique cutoff value.

In each community obtained using CFinder, we performed a pathways enrichment analysis followed by a ranking of all pathways. This ranking or scoring was done as follow: if $ConsenScore_i^k$ is the $ConsenScore_i$ of the gene “i” in the community “k” then:

- 1) Each community “k” was weighted as: $W_k = \sum ConsenScore_i^k / N_k$, where N_k is the number of communities.
- 2) Each pathway “m” was weighted as: $PathRankScore_m = \sum W_k^m / N_k^m$, where W_k^m is the weight (W_k) of each communities connected with the pathway “m” and N_k^m is the number of communities connected with the pathway “m”.
- 3) A second weight was given to the pathway “m” ($PathGeneScore_m$) considering all the genes involved in the pathway as: $PathGeneScore_m = \sqrt{\langle ConsenScore_i^m \rangle \frac{n_m}{N_m}}$, where N_m is the total number of genes in the pathway “m” while n_m is the number of those genes which are also found in the protein-protein interaction network. $\langle ConsenScore_i^m \rangle$ is the average of the $ConsenScore_i$ of all genes presents in the pathway “m”.
- 4) The final score associated with the pathway “m” ($PathScore_m$) is calculated as the geometrical mean between $PathGeneScore_m$ and the normalized $PathRankScore_m$.

Microarray analysis

A total of five studies were considered in microarray data integration and were named as: A1 [7], A2 [6], A3

[14], A4 [8] and A5 [10]. In each study we extracted the significant up-regulated/down-regulated genes considering the procedure of each author in the correspondent articles. In any case was considered the fold-expression as significant cutoff but the adjusted p -value reported by the authors. The criterion was an adjusted p -value < 0.05 . The strategies in the reported articles considering: microarrays integration, gene expression correction and annotations where different (we will discuss more about it in results section and a brief description can be found in Additional file 2). The adjusted p -values were used to create a ranking of genes in each study followed by independent normalization.

We could go through a meta-analysis cross-normalization approach as in [10, 33]. However, because different strategies are possible to accomplish this analysis leading to different results we choose to consider each study separately. In each study (“j”) a particular up-regulated or down-regulated gene (“i”) will have a normalized score according to ranking ($GeneS_{i,j}$ the normalized score of the gene “i” in the study “j”). The consensus scoring of each gene in the microarray data was carried out similarly to consensus prioritization strategy. This means, the final score of each gene was calculated as

$GeneAS_i = \sqrt{\left(\frac{N_{array_i} - 1}{5 - 1}\right) \left(\frac{1}{j} \sum_j GeneS_{i,j}\right)}$ where, N_{array_i} correspond with the number of studies reporting the gene “i”. Combining all genes in the selected studies we found 1944 genes: 916 always reported up-regulated, 1013 always reported as down-regulated and 13 genes with ambiguous expression. The full list of genes is presented in Additional file 2 as well as the calculated scores. This final score ($GeneAS_i$) will have a double meaning 1) inter-studies agreement and 2) the measure of the statistical significance of the gene in the study. Therefore, highest values of the score imply that the gene was identifying in several studies and also with highest statistical significance.

Results

Consensus prioritization

The detections of pathogenic genes in all methods are presented in Table 1. As we can notice not all methods are capable to identify the 35 proposed pathogenic genes.

Consensus strategy identify the entire G2 set in the first 1% of the final gene list (>18,000 genes) and in all cases remain as the method with higher identification of pathogenic genes. Very close to consensus strategy approach is the MetaRanker [28] method. The identification of the pathogenic genes is important but even more relevant is the early recognition ability.

The average rank of the studied genes is lower in consensus strategy than in other methods (Table 2) used independently. The average rank of the detected genes is not properly speaking a measure of early recognition.

Table 1 Identification (in %) of pathogenic genes in each approach

Methods	1%			5%			10%			20%		
	G1	G2	G1,2									
BioCarta	0,00	0,00	0,00	0,00	7,69	2,86	3,70	23,08	8,57	3,70	23,08	8,57
Candid	14,81	15,38	14,29	25,93	61,54	34,29	29,63	69,23	37,14	44,44	84,62	54,29
GLAUG4	3,70	0,00	2,86	14,81	15,38	14,29	18,52	53,85	28,57	22,22	76,92	37,14
PlySearch	0,00	0,00	0,00	0,00	7,69	6,25	6,25	15,38	12,50	12,50	23,08	18,75
CIPHRE	0,00	0,00	0,00	0,00	7,69	2,86	0,00	15,38	5,71	0,00	15,38	5,71
Guildify	14,81	23,08	14,29	18,52	38,46	22,86	25,93	53,85	34,29	44,44	69,23	51,43
DISGENET	3,70	15,38	5,71	7,41	30,77	14,29	11,11	38,46	20,00	29,63	92,31	45,71
Geneprospector	3,70	30,77	11,43	22,22	84,62	34,29	22,22	100,00	40,00	33,33	100,00	48,57
GENIE	7,41	0,00	5,71	14,81	46,15	25,71	25,93	76,92	40,00	37,04	84,62	48,57
SPNS3D	7,41	7,69	5,71	14,81	15,38	11,43	14,81	30,77	17,14	29,63	53,85	34,29
GeneDistiller	0,00	0,00	0,00	0,00	7,69	6,25	6,25	7,69	12,50	12,50	15,38	18,75
MetaRanker	44,44	92,31	54,29	62,96	92,31	68,57	74,07	92,31	77,14	88,89	92,31	88,57
Consensus	51,85	100,00	62,86	74,07	100,00	80,00	85,19	100,00	88,57	88,89	100,00	91,43

However, intuitively will means that consensus strategy early detect the pathogenic genes (identified in G1 and G2 groups). The MetaRanker is one more time the closer strategy. Even when these two previous analyses could indicate that the consensus strategy prioritize better the pathogenic genes, we additionally calculate several indexes directly related with the evaluation of early enrichment (Table 3). Because MetaRanker is the method with closer results, the early enrichment analysis was only performed comparing consensus and MetaRanker strategies.

The indexes related with the early enrichment clearly state that consensus strategy over perform the result of MetaRanker in pathogenic genes detection locating more genes with a significant lower rank. We compare

the rank of the pathogenic genes between the two methods for G1, G2 and G1,2 using signed Wilconson test. The p -value was lower than 0.01 in the three groups indicating statically significant differences in the ranking obtained by the two methods.

Previous calculations are based on predefined genes in G1 and G2 groups. In order to explore the consistency of our results by changing those genes, we performed a bootstrap sampling as follow:

- We remove 5 random genes form the 35 “pathogenic genes” (around 14%) and evaluate the median rank of the remaining ones in both: consensus and MetaRanker

Table 2 Average rank of identified pathogenic genes in each method

Methods	1%			5%			10%			20%		
	G1	G2	G1,2	G1	G2	G1,2	G1	G2	G1,2	G1	G2	G1,2
BioCarta					4,0	4,0	7,0	5,7	5,7	7,0	5,7	5,7
Candid	18,8	9,0	17,2	44,1	70,9	58,3	71,3	92,0	73,8	180,7	157,9	182,5
GLAUG4	1		1	3,0	5,0	3,2	4,8	8,3	6,4	6,5	11,2	9,1
PlySearch	0	0	0		1,0	1,0	2,0	1,5	1,5	3,0	2,3	2,3
CIPHRE					15,0	15,0		28,5	28,5		28,5	28,5
Guildify	36,0	41,0	29,0	57,0	155,6	117,6	300,1	324,0	353,9	1010,9	648,9	864,1
DISGENET	2,0	1,5	1,5	3,0	2,8	3,0	6,3	4,6	5,7	15,5	13,4	13,8
Geneprospector	4,0	2,5	2,5	11,2	7,7	7,7	11,2	9,8	9,6	24,0	9,8	16,7
GENIE	2,0		2,0	5,8	7,7	6,4	11,0	12,3	10,9	19,2	13,8	15,8
SPNS3D	2,0	3,0	2,0	4,0	5,5	4,0	4,0	8,0	6,2	12,9	13,4	13,9
GeneDistiller	0,0	0,0	0,0		1,0	1,0	1,0	1,0	1,0	1,5	1,5	1,3
MetaRanker	45,8	43,7	44,9	143,8	43,7	114,5	297,3	43,7	231,4	648,6	43,7	511,9
Consensus	36,4	15,8	28,0	118,2	15,8	88,3	272,6	15,8	205,7	372,0	15,8	282,4

Table 3 Initial enrichment indexes for the MetaRanker and the Consensus strategy

Indexes	MetaRanker			Consensus		
	G1	G2	G1,2	G1	G2	G1,2
MEDIAN RANK	246 (6–12,345)	23 (3–3709)	154 (3–12,345)	114 (2–9218)	10 (1–59)	46 (1–9218)
AUC	0,927	0,982	0,938	0,944	1	0,957
EF_1%	44,533	92,491	54,394	51,993	100,272	63,028
EF_5%	12,604	18,478	13,726	14,823	20,011	16,009
EF_10%	7,41	9234	7717	8519	10	8857
EF_20%	4445	4616	4429	4444	5	4571
RIE_1%	38,836	75,295	46,407	48,1	92,346	58,435
RIE_5%	11,807	17,656	12,941	13,806	19,673	15,158
RIE_10%	6912	9117	7324	7676	9918	8191
RIE_20%	3971	4731	4,11	4216	5013	4399
BEDROC_1%	0,418	0,78	0,51	0,517	0,956	0,642
BEDROC_5%	0,599	0,889	0,66	0,7	0,991	0,772
BEDROC_10%	0,696	0,915	0,739	0,773	0,995	0,827
BEDROC_20%	0,79	0,941	0,819	0,84	0,998	0,877

b) We repeat the previous step 1000 times, each time selecting a new set of 5 random genes.

The density distribution (using Gaussian kernel of function “density” in R) of the 1000 values in both methods is presented in the Fig. 1.

As we can noticed, consensus strategies compared to MetaRanker provides more frequently the lower rank for the genes in agreement with our previous results in Table 3 after evaluation in 1000 samples modifications.

Enrichment analysis of preeclampsia related genes and protein-protein interaction network

Using G1 and G2 unique genes ($n = 35$), we can notice (Table 1) that consensus strategy already identify the 89% in the 10% of the data, this means that the 89% of the 35 genes are in the initial 1800 genes obtained from prioritization. This is a very big number; therefore a

strategy for a rational cutoff was designed (see Methods). The implementation of I_i considering the true positive and false positive ratio could be used as a rational cutoff to reduce the amount of genes. This procedure is showed in Fig. 2.

The maximal (Fig. 2) value of I_i is 0.76085 and correspond with a ranking value of 476, therefore the reduced list for PE comprise the first 476 genes. The entire gene list as well as their scores and ranking can be found in Additional file 3. In the 476 genes there are 30 of 35 predefined pathogenic genes.

The enrichment analysis of biological processes in these genes results into more than 500 terms with an adjusted p -value < 0.01 (considering FDR) (Additional file 4). In order to simplify this list we used Revigo [37] to calculate the frequencies of the gene ontology terms. We only consider those terms with a frequency lower than 0.01% (full list of terms can be found in Additional file 4). With

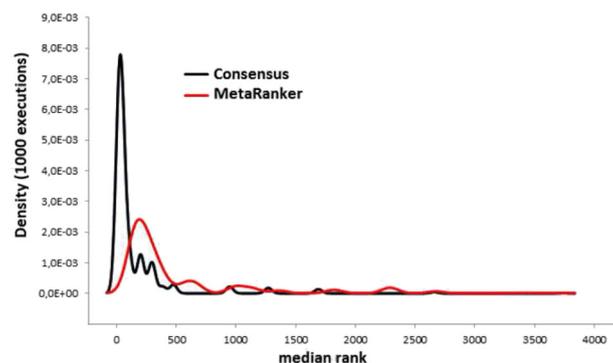


Fig. 1 Average ranking distribution in consensus and MetaRanker strategies in 1000 generations randomly removing the 14% of the pathogenic genes (G1,2) each time

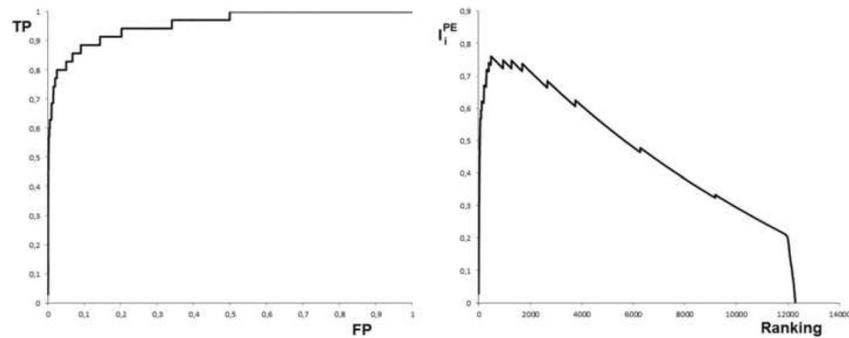


Fig. 2 Left) ROC curve obtained with prioritized genes for PE and the proposed pathogenic list. Right) Variation of I_i with respect to genes ranking. The maximal value of I_i is the 0.76085 and correspond with a ranking value of 476

this consideration the number of terms remains high so only some of the initial ones (more relevant) are presented in Table 4.

Similarly, the enrichment analysis of metabolic pathways is presented in Table 5 using to main databases: KEGG and Reactome.

The pathways presented in Table 5 are only a partial list but it is entirely presented for Reactome and KEGG in Additional file 4.

The biological processes and enriched pathways are consistent between them and also with the scientific knowledge about PE, however would be hard to establish

Table 4 Some of the more specific biological process obtained by enrichment analysis in PE genes

BP ID	Name	Frequency	log ₁₀ p-value
GO:0008217	regulation of blood pressure	0,01%	-33,0101
GO:0032496	response to lipopolysaccharide	0,01%	-16,8268
GO:0030193	regulation of blood coagulation	0,01%	-16,684
GO:0050818	regulation of coagulation	0,01%	-16,6635
GO:0043434	response to peptide hormone	0,01%	-15,6819
GO:0048660	regulation of smooth muscle cell proliferation	0,00%	-15,3242
GO:0032868	response to insulin	0,01%	-13,5114
GO:0045765	regulation of angiogenesis	0,01%	-13,1612
GO:0070663	regulation of leukocyte proliferation	0,01%	-12,3862
GO:0031960	response to corticosteroid	0,00%	-12,214
GO:0045428	regulation of nitric oxide biosynthetic process	0,00%	-11,8447
GO:0043627	response to estrogen	0,01%	-11,8447
GO:0050670	regulation of lymphocyte proliferation	0,01%	-11,3468
GO:0007568	aging	0,01%	-11,3468
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	0,01%	-11,1844
GO:0003073	regulation of systemic arterial blood pressure	0,01%	-11,0177
GO:0051384	response to glucocorticoid	0,00%	-10,9101
GO:0010743	regulation of macrophage derived foam cell differentiation	0,00%	-10,8962
GO:0050729	positive regulation of inflammatory response	0,01%	-10,8962
GO:0050886	endocrine process	0,00%	-10
GO:0031099	regeneration	0,01%	-9,9245
GO:0051341	regulation of oxidoreductase activity	0,00%	-9,8996
GO:0019229	regulation of vasoconstriction	0,00%	-9,8761
GO:0042035	regulation of cytokine biosynthetic process	0,01%	-9284
GO:0019218	regulation of steroid metabolic process	0,01%	-8,1221

Table 5 Pathways enrichment analysis using Reactome and KEGG databases

Pathway Name (KEGG)	% Genes	p-value
Cytokine-cytokine receptor interaction	16,033,7553	1,22E-26
Complement and coagulation cascades	7,59,493,671	1,64E-20
Graft-versus-host disease	4,85,232,068	1,33E-13
Allograft rejection	4,21,940,928	1,28E-10
Focal adhesion	9,28,270,042	4,28E-09
Type I diabetes mellitus	4,21,940,928	4,58E-09
Antigen processing and presentation	5,69,620,253	1,25E-08
Hematopoietic cell lineage	5,48,523,207	1,95E-07
Jak-STAT signaling pathway	7,17,299,578	1,91E-06
Renin-angiotensin system	2,32,067,511	2,63E-05
TGF-beta signaling pathway	4,64,135,021	2,13E-04
Adipocytokine signaling pathway	4,00843882	2,91E-04
Endocytosis	6,96,202,532	5,37E-04
Natural killer cell mediated cytotoxicity	5,69,620,253	6,70E-04

Pathway Name (Reactome)	% Genes	FDR
Hemostasis	15,8,227,848	7,21E-29
Signaling in Immune system	12,2,362,869	7,53E-11
Signaling by PDGF	5,06329114	1,73E-08
Signaling by VEGF	2,10,970,464	2,00E-06
Integrin cell surface interactions	4,85,232,068	2,05E-05

some kind of relevance between them without further consideration. In this way we carried on a network analysis.

With the indicated cutoff of 0.9, the final protein-protein interaction network has 417 nodes, corresponding with the 87.6% of the initial genes (476). The S^k index (as proposed in the Methods section to identify a rational k-clique number) will rich a minimum either by

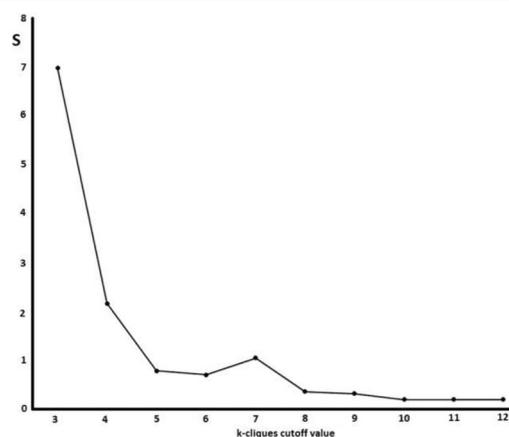
an increment in the number of communities and/or by an increment in the similarity between the mean and median values of the number of genes in all communities. We can notice (Fig. 3) that the desired values will be between 8 and 10. The number of communities for $k = 8$ is 16 compared to 9 and 5 for $k = 9$ and 10 respectively. Considering that in each community several biological analyses will be carried on, 16 communities will be difficult to study. Additionally in $k = 8$, one of the communities have almost twofold the number of genes with respect the remaining communities. For this reason we select the $k = 9$ in our analysis (Fig. 4. Left).

Each community can be weighted considering the *ConsenScore* of each gene in the community (see Table 6). Additionally we also included the number of pathogenic genes present in the community.

Communities 2 and 6 could be considered as the more relevant showing. However could be useful the prioritization of the metabolic pathways by an enrichment analysis in each community (*full list presented in Additional file 4*) and weighted as presented in the Methods section.

Microarray data integration

From the 1944 genes collected in microarrays data only 80 are present in our 476 obtained in consensus prioritization representing only a 4%. The worst gene overlapping with respect to other microarrays is with the study A1. The A2, A3, A4, A5 conserve 40 genes in common but drastically reduced to 2 by adding A1 (Fig. 5 Left). The agreement between selected microarray studies is not good in terms of genes identifications as we can see in the Venn diagrams (Fig. 5, Left). It is a direct consequence of the differences in initial microarray data and processing strategies (presented in Additional file 2). The study A1 is the only one with any meta-analysis strategy. Both A2 and A3 carried out a meta-analysis, while A4 and A5 go

**Fig. 3** Values of S^k with respect to each k-clique cutoff value

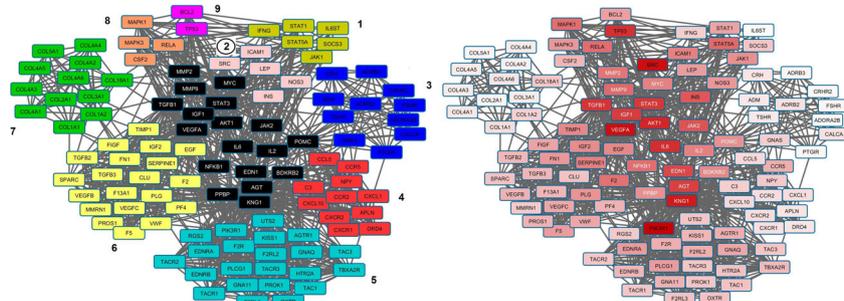


Fig. 4 *Left*). Community analysis for k-cliques = 9. *Black* nodes represent genes which are parts of several communities. The rest of the colors correspond with the 9 communities obtained. *Right*) Gradient connectivity degree distribution (min = 9 with *white* color and max = 85 with *red* color and indicated by PIK3R1 gene)

specifically through cross-platform normalization. The differences between both strategies in microarray data integration had being explored previously [42]. Actually A2 and A3 share 111 genes and similarly A4 and A5 share 237 genes. This gene space is reduced 40 genes when all four studies are combined. Moreover as we can see in Additional file 2, A4 and A5 share a number of similarities regarding initial microarray data.

Analyzing the amount of genes that each study independently shares with the initial 476 prioritized genes (Fig. 5. Right) we found that: A1 ($n = 12, 3.8\%$), A2 ($n = 30, 7.7\%$), A3 ($n = 30, 7.9\%$), A4 ($n = 53, 4.2\%$) and A5 ($n = 26, 7.5\%$). The result indicates that the

methodology of Moslehi et al. [14] will represent better our prioritized genes (even when very close to A2 and A5). Moreover considering the average consensus score of those genes we found: A1 (0.396), A2 (0.561), A3 (0.597), A4 (0.111) and A5 (0.540). This average scoring also suggests that the work of Moslehi et al. [14] also cover better ranked genes (even not so distant of A2 and A5). These values will be discussed later.

There are a total of 41 up-regulated and 39 down-regulated genes commonly found between all integrated genes in microarray data and the 476 already prioritized genes (a total of 80 genes). The up-regulated are: VEGFA, FLT1, STOX1, SERPINE1, LEP, INHA, INHBA,

Table 6 Communities membership and scores

Communities	Genes	Average ConsenScore _i	Average Rank	Average Degree	N Pathogenic	N Pathways
2	TGFB1, SRC,IGF1,IL6, INS, LEP, NOS3, AKT1, ICAM1, MMP2, STAT3, VEGFA, EDN1, MMP9	0,995	89,07	55.21	5	8
6	TGFB1, TGFB3, EGF, VWF, IGF1, F2, KNG1, PPBP, SERPINE1, TIMP1, FN1, PLG, VEGFA, IGF2, CLU, F13A1, FIGF, MMRN1, PF4, SPARC, VEGFB, VEGFC, F5, TGFB2, PROS1	0,991	171,28	38.24	5	7
9	TGFB1, IL6, NFKB1, TP53, AKT1, MMP2, BCL2, MYC, MMP9	0,989	194,56	52.33	2	6
1	IL6, IL2, STAT3, IFNG, STAT5A, JAK2, JAK1, SOCS3, STAT1, IL6ST	0,989	210,2	39.60	1	4
8	MAPK1, MAPK3, NFKB1, IL2, STAT3, RELA, CSF2, JAK2, MYC	0,989	210,67	44.89	0	14
5	AGT, AGTR1, BDKRB2, EDNRA, F2R, F2RL2, F2RL3, GNA11, GNAQ, KNG1, OXTR, PIK3R1, PLCG1, PROK1, RGS2, TAC1, TAC3, TACR1, TACR3, UTS2, EDN1, EDNRB, HTR2A, KISS1, TACR2, TBXA2R	0,987	231,04	33.88	4	7
3	ADM, CRH, POMC, ADORA2B, PTGIR, TSHR, ADRB2, ADRB3, CALCA, CRHR2, GNAS, FSHR	0,987	243,92	16.50	1	3
4	POMC, CCR5, AGT, APLN, BDKRB2, C3, CCL5, CCR2, CXCL10, CXCR1, CXCR2, DRD4, IL8, KNG1, NPY, PPBP, CXCL1	0,985	273,25	28.75	1	5
7	COL18A1, COL1A1, COL1A2, COL3A1, COL4A5, COL2A1, COL4A1, COL4A2, COL4A6, COL5A1, COL4A3, COL4A4	0,983	304,5	15.58	0	2

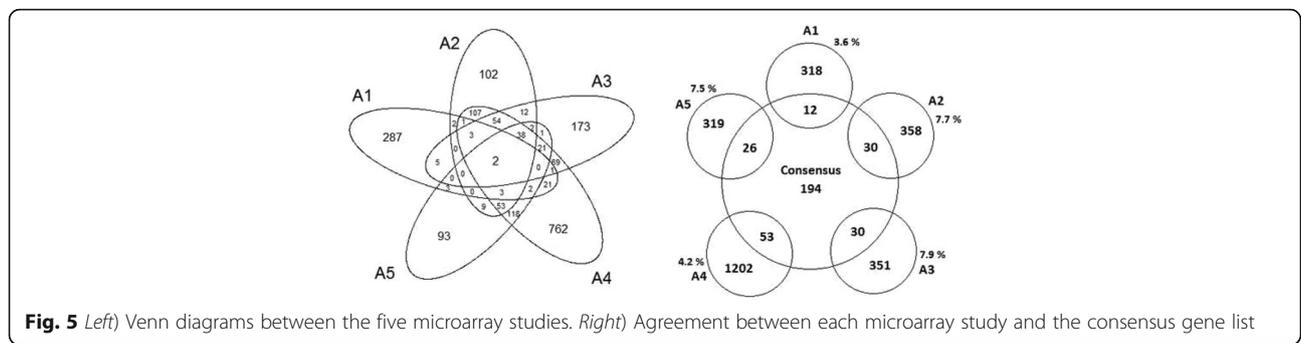


Fig. 5 Left) Venn diagrams between the five microarray studies. Right) Agreement between each microarray study and the consensus gene list

ENG, HMOX1, VWF, TGFB1, TFPI, ADAM12, CRH, PAPP2, VEGFC, CP, MMP14, FN1, SERPINA3, SIGLEC6, ACE2, PREP, FABP4, EGFR, FSTL3, IL6ST, VDR, IGFBP5, MMP15, ITGA5, TRIM24, CGA, MET, DUSP1, MIF, TAPBP, NR1H2, MMP11, HPN, GLRX and the down-regulated are: ACVRL1, ADRB3, AGTR1, ANGPT1, CD4, CD14, COL1A1, COL1A2, F5, F13A1, FCER1G, FGF2, GHR, CFH, HGF, HSD11B2, CFI, IGF1, IGFBP7, IL10RA, IDO1, JAK1, KLRD1, MMP1, NEDD4, ENPP1, PLAUR, MAPK1, CCL2, SOD1, SPP1, TGFBR3, THBS1, TLR4, VCAM1, APLN, HGS, ROCK2, PLAC1. From these 80 genes 34 (42.5%) were located with a ranking less than 180 (around the first 1% of the list) in the consensus strategy prioritization.

Comparing the scores of consensus strategy and scores obtained from the microarrays studies (Fig. 6) we can arrive to some interesting results. Moreover, from these 80 genes 72 are also present in the 417 forming the interaction network and 19 are also part of some community. We can evaluate the contribution of these 19 genes in each community using the average $GeneAS_i$ of the genes which belong to a particular community in a similar way as we did previously (Table 6). The corresponding weights for

each community are: 1 (0.062), 2(0.140), 3(0.072), 4(0.033), 5(0.014), 6(0.132), 7(0.054), 8(0.031) and 9(0.048). These weights also confirm that communities 2 and 6 could be the more relevant as previously presented.

Integrated metabolic network

Using RSpider [38] from the 476 genes only 272 were mapped to a reference global network obtaining three significant models (Table 8).

The p -value indicates the probability for a random gene/protein list to have a maximal connected component of the same or larger size. This p -value is computed by Monte Carlo simulation as described in [38]. Beside this statistical analysis, we should also consider that in the initial data (476) there are 80 genes also matching with microarray data while in the smallest network 23 of 98 are also present in the microarray information. This enrichment is statistically different (p -value = 0.036) compared to random gene extraction. The 23 genes are: ANGPT1, COL1A1, COL1A2, F5, F13A1, FCER1G, FLT1, FN1, IGF1, IGFBP5, ITGA5, JAK1, MET, MMP1, SERPINE1, PLAUR, SPP1, TFPI, THBS1, VEGFA, VEGFC, VWF and PAPP2. The network associated with Model 1

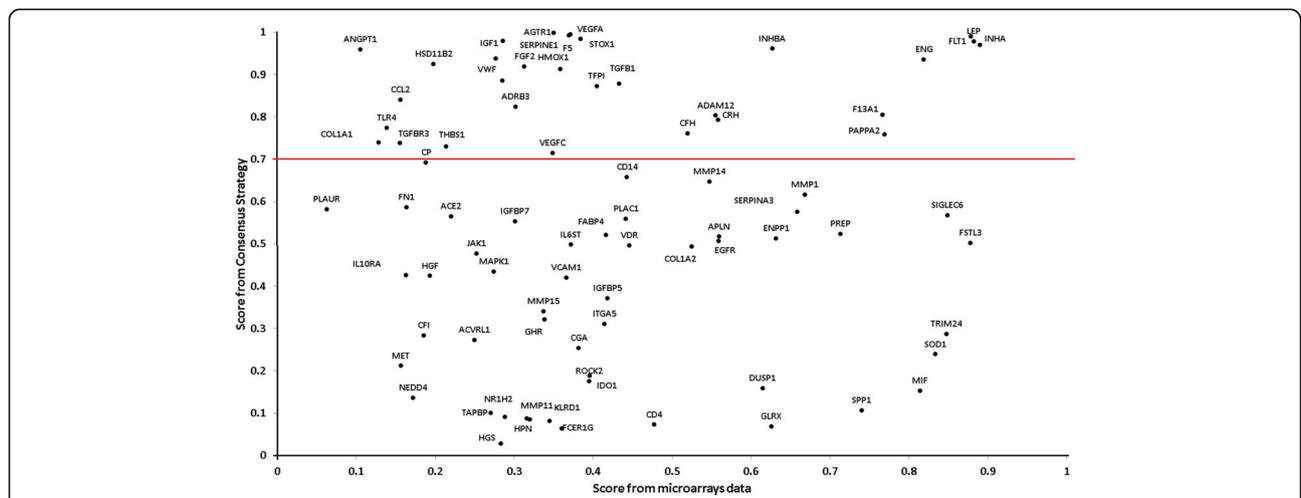


Fig. 6 Relationship between the score obtained from microarray data and the consensual strategy prioritization. The red line indicates and scores in consensus prioritization equal to 0.7

is presented in Fig. 7. The network in Model 3 of Table 8 is presented in Additional file 5.

The expanded integrated metabolic network (Model 3) allows the entrance of 114 genes in order to bring connections between initial genes. However, it also incorporates 32 compounds that also act as connectors. These compounds obtained from the integrated network are presented in Table 8, some of them could be very generic like “fatty acids” or could be very specific as “serotonin” or “L-Homocysteine”. These compounds can be easily grouped mainly in lipids, steroids, amino acids, and purine metabolisms and will be further explored in the discussion.

Discussion

Consensus prioritization and enrichment analysis

Our results confirm that the consensus strategy actually improve the detection and prioritization of pathogenic genes. Application of early recognition measures are important and should be considered together with identification capabilities. The ability to rank the relevant genes on the top of a long prioritized list will directly reduce cost of experimental validation. Previous authors had being probed that consensual strategy in prioritization improve the detection of genes related with specific pathology [15, 16, 33]. However, we are proving here for the first time that the consensus strategy also improves the early enrichment ability of genes related with pathogenesis (at least in PE).

Any study in gene-disease association is intrinsically focused into pathogenesis discovery. During this process

obviously some relations could be established and not necessarily because of pathogenesis but a secondary modification (the experimental design will be directly related with this type of result). If several prioritization strategies are combined, then, the possibility of removing noisy relationships (in pathogenesis terms) increases as well as the agreement in relevant genes.

The biological processes as well as metabolic pathways enrichment analysis lead us to already expected information. Some of the biological processes, like those related to blood pressure or vasoconstrictions have a direct association with PE clinical development. Biological processes associated with inflammation, angiogenesis, cytokine, immune system and hormones regulation could also be associated with PE clinical manifestation or even pathogenesis [6, 7, 43, 44] and there are also well related with the metabolic pathways enrichment results (Table 5). The pathway analysis also reflects a good agreement with previous works. The cytokine pathway, VEGF and PDGF signaling, immune system and even some of the cancer related pathways were previously reported by other authors [6, 7, 9, 14, 44, 45]. Signaling pathways in general, are highly relevant as well as several routes connected with cancer (see Additional file 4) which also agree with the previous studies [14, 46, 47].

Protein-protein interaction network, communality analysis and microarrays integration

The enrichment analysis can be helpful. However it is hard to establish a ranking of the pathways according to their implications in pathogenesis without further analysis. It is

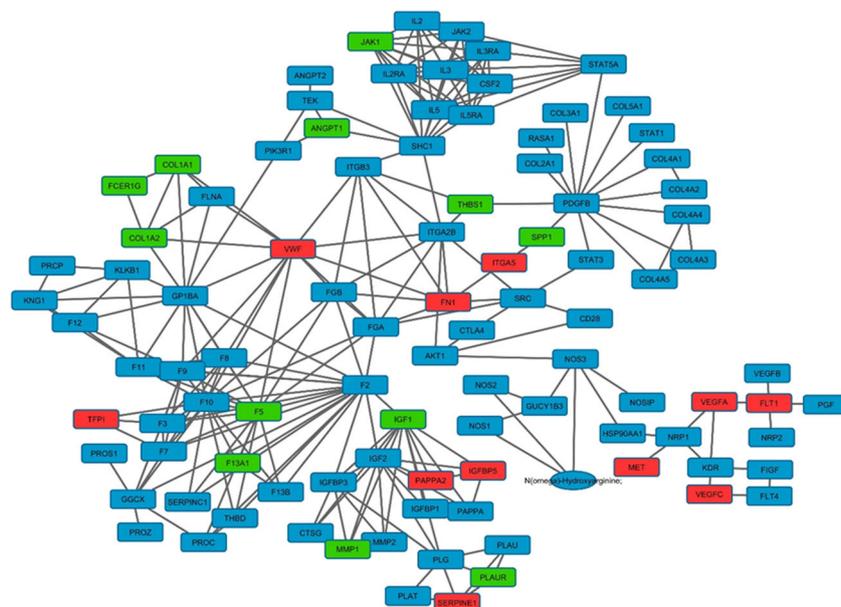


Fig. 7 Integrated metabolic network with 98 genes colored according to our microarray data. The color are: green, red and blue, indicating down-regulated, up-regulated and no information from microarray respectively

the main reason to combine the analysis of the protein-protein interaction network. The entire network contains 417 nodes but only 111 are part of some community. The network with 417 already comprises 29 of the 35 predefined pathogenic genes. The sub-network containing only genes which belong to some community have 12 of 29 predefined pathogenic genes. Moreover only 3 genes (HADHA, IDO1 and HLA-G) of the remaining 17 genes are not directly connected with the sub-network. On the other hand, the average degree of the pathogenic genes is 23.6 which is statistically significant higher than non-pathogenic genes (14.2) at p -value <0.05 . This result indicates that the node degree could be associated with pathogenesis in this network. The black colored nodes represent those genes that are present in more than one community and therefore are usually those with higher connectivity degree as we can also see in the Fig. 3 Right (and Table 6).

The top 20 genes with highest connectivity degree are: PIK3R1, SRC, VEGFA, KNG1, AKT1, IL6, TP53, TGFB1, STAT3, IGF1, AGT, EDN1, JAK2, INS, EGFR, SHC1, MAPK8, MMP9, STAT5A and MAPK1. The majority of them are located between communities (black colored) and only 3 (EGFR, SHC1 and MAPK8) were not identified as member of any community. The community's analysis (Table 6) indicates that communities 2 and 6 could be considered as the more relevant showing: 1) the highest scores; 2) the minimal average ranking and 3) both include the major number of pathogenic genes. In terms of connectivity degree the community 2 have the greater value instead community 6 which have a middle one. Looking at genes in the community 2 we can clearly identify elements of the VEGF signaling and also NOS metabolism through AKT1 and generally a core of possible mechanism well established in PE that will be discussed later. Moreover, the prioritization of the metabolic pathways shown that VEGF signaling pathway is not only the most relevant pathway (Table 7) but also it is exclusively enriched in community 2.

Actually we can notice that tops pathways primarily involve communities 2 and 6. It indicates that those communities as well as their genes are highly relevant in PE. Additionally, the community number 5 is exclusively related with the Renin-angiotensin system and considering that it is also enriched in neuroactive ligand-receptor interaction and vascular smooth muscle contraction, we can suspect that this community has a strong connection with the hypertensive disorder. Interestingly community 8 have the major number of associated pathways. However, most of them are related with signaling pathways like TGF-beta signaling. The enrichment in signaling processes could indicate that it is probably a central group of genes acting as connectors between several metabolic processes and therefore would be relevant to comprehend PE heterogeneity.

Following the importance of community 2 and 6, the major pathways ordered by relevance connected to both communities are: VEGF signaling pathway, mTOR signaling pathway, Adipocytokine signaling pathway, Intestinal immune network for IgA production, Leukocyte transendothelial migration, Progesterone-mediated oocyte maturation, Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, complement and coagulation cascades, TGF-beta signaling pathway, focal adhesion and regulation of actin cytoskeleton.

In order to explore our results using additional experimental information we included the microarrays analysis. The worst genes overlapping with respect to prioritized list is with the study A1 (Fig. 5), while the other four studies show more consistent results. The reason for this difference in A1 is mainly because is the only study that is not a meta-analysis. We included it because it is the largest independent study.

Our result indicates that the study of Moslehi et al. [14] identify more common genes and also better ranked in our consensus strategy. Both studies using meta-analysis (A2 and A3) shown a better agreement with consensus than A4 and A5. The A5 is better than A4, and similar to A2 and A3 probably because the use of *combat* [48] and an increased number of arrays. Also, A2 is the study that carried out the largest microarray integration related with PE. The use of *combat* in cross-platform normalization had being favored in term of clinical and biological meaning agreement [42]. The differences in the percentage of genes (Fig. 5 Right) shared with consensus strategy is really small comparing A3 with A2 and A5 studies. The A3 study also comprise some similarities with A4 and A5 regarding initial microarray data. However, A3 study exclusively considers meta-analysis in microarrays of early-onset preeclampsia. It is a very important difference regarding other studies. We had probed that consensus prioritization actually improve pathogenic early recognition and we also known that genes involved in early onset preeclampsia are probably closer to pathogenesis than late-onset preeclampsia. This could explain why the A3 have the highest average score with our prioritized consensus list. Therefore it is a logical result considering the previous analysis and also an indirect validation of our consensus strategy. The A5 and A4, consider similar microarrays than A3 but including other do not exclusively related with early-onset preeclampsia.

Regarding all differences between microarrays studies, we should remember that all genes extracted from microarray data were statistically significant up or down-regulated in each corresponding study. Moreover we carried out a complete integration of the gene space between all microarrays data considered, so, to our effects we didn't exclude any genes because be part or not of a particular study. Therefore, this disparity in microarrays studies could only affect the *GeneAS_i* scoring. The score should be interpreted as a commitment between agreement across methods and

Table 7 Pathways enrichment analysis in communities and their associated weights

Pathways	<i>PathRankScore_k</i>	N Community	<i>PathGeneScore_m</i>	<i>PathScore_m</i>	Community
VEGF signaling pathway	89,07	1	0,422	0,547	2
mTOR signaling pathway	130,18	2	0,445	0,505	2,6
Adipocytokine signaling pathway	169,98	3	0,518	0,479	1,2,8
Intestinal immune network for IgA production	171,28	1	0,497	0,467	6
Leukocyte transendothelial migration	89,07	1	0,303	0,463	2
Progesterone-mediated oocyte maturation	89,07	1	0,301	0,462	2
Cytokine-cytokine receptor interaction	185,95	4	0,529	0,455	1,2,4,6
Jak-STAT signaling pathway	176,12	4	0,468	0,445	1,2,8,9
Renin-angiotensin system	231,04	1	0,804	0,443	5
MAPK signaling pathway	149,87	2	0,378	0,439	8,9
Complement and coagulation cascades	225,19	3	0,708	0,432	4,5,6
TGF-beta signaling pathway	190,97	2	0,486	0,427	6,8
Focal adhesion	188,28	3	0,464	0,422	2,6,7
Apoptosis	194,56	1	0,431	0,396	9
Regulation of actin cytoskeleton	171,28	1	0,326	0,378	6
Natural killer cell mediated cytotoxicity	210,67	1	0,453	0,375	8
ErbB signaling pathway	210,67	1	0,451	0,374	8
Fc epsilon RI signaling pathway	210,67	1	0,450	0,374	8
T cell receptor signaling pathway	210,67	1	0,436	0,368	8
Neurotrophin signaling pathway	202,61	2	0,339	0,338	8,9
Toll-like receptor signaling pathway	226,16	3	0,432	0,335	4,8,9
NOD-like receptor signaling pathway	241,96	2	0,510	0,326	4,8
Dorso-ventral axis formation	210,67	1	0,331	0,321	8
B cell receptor signaling pathway	210,67	1	0,328	0,319	8
Cell cycle	194,56	1	0,252	0,303	9
Chemokine signaling pathway	231,37	3	0,370	0,300	1,4,8
Gap junction	231,04	1	0,351	0,293	5
Calcium signaling pathway	237,48	2	0,363	0,284	3,5
Neuroactive ligand-receptor interaction	237,48	2	0,343	0,276	3,5
Vascular smooth muscle contraction	237,48	2	0,327	0,270	3,5
Melanogenesis	231,04	1	0,281	0,262	5
ECM-receptor interaction	304,50	1	0,478	0,040	7

their statistical significance. Even when is reasonable to assume that a gene with a simultaneously high agreement and high statistical significance could be very important (i.e. LEP, FLT1, INHA) (Fig. 6), also the condition of high agreement with low statistically significance is equally relevant (because actually leads to highest score). In other words, a statistical significance don't necessarily means that the gene is more relevant to the disease than any other with a reduced but significant change.

Previously we presented evidence indicating that consensus prioritization is capable to identify genes with high pathogenic probabilities in the first portion of the data. It is clearly presented in Fig. 6 with VEGFA,

AGTR1, F5 and TGFB1 which are well related with pathogenesis [49–55]; however, the score obtained from microarray data is relatively low in these cases (less than 0.5). Considering a high cutoff value (i.e. >0.7) we can identify: LEP, FLT1, INHA, ENG, PAPP2, and CRH. There are sufficient evidences to associate these genes with PE pathogenesis or clinical manifestation [14, 52, 53, 56–60]. Our calculations also indicate that communities 6 and 2 are those with highest enrichment of genes coming from microarray data confirming our previous results using the network and consensus prioritization (Table 6). This consistency support that the prioritization strategy is actually pointing us in the correct direction

and also justify the idea that the associated pathways could also be highly relevant.

Metabolic involvement

In all previous analysis, the VEGF signaling pathways had being selected as the most relevant in PE. This pathway is presented (Fig. 6) with the genes: VEGFA, VEGFB, VEGFC, FLT1, KDR, FLT4, PGE, NRP2 and NRP1. These genes are directly connected with arginine (NOS1, NOS2 and NOS3) and nitric oxide metabolisms (NOSIP and HSP90AA1). Considering the previous results in communities and pathways enrichment analysis we should conclude that this processes will be actually the most significant for PE pathogenesis. Interestingly the involvement of VEGF, FLT1 and several elements in arginine metabolism, including NO production, was proposed in [61] as the primary mechanism in placenta leading to PE.

In the protein-protein interaction network analysis we can noticed that genes like VEGFA, NOS3, SRC and AKT1 are highly connected (community 2 in Table 6 and Fig. 4) but in the integrated metabolic network (Model 1, Fig. 7) their connectivity is not so elevated (or even in the extended integrated metabolic network showed in Additional file 5). The reason for these differences is a direct consequent of the pathway representation in KEGG and Reactome. For example, VEGF mediates the ezrin/calpain/PI3K/Akt pathway-dependent stimulation of NOS3 phosphorylation leading to Ca²⁺ independent NO generation [62–64]. This connection will be reflected in the PPI as an edge between VEGFA and NOS3 or even between VEGFA and PIK3R1 (distant nodes in Fig. 6) and these both interactions will not be seen in the Fig. 7. This is why we should additionally use the Model 3 for integrated pathway (presented in Additional file 5) and consider the pathway in Fig. 7 as the simplest representation of the biological meaning of genes involved in PE. Form the metabolic integrated network (Fig. 6) we can clearly identify some well-known mechanisms related with VEGF and PE and other relevant effects that will be discussed further.

The increment in FLT1 production (noticed by the microarray data) could lead to the increment in the soluble-Flt1 rescuing the extracellular VEGFA [65, 66]. Therefore the increment in VEGFA expression in placenta could be a compensatory response to restore normal angiogenesis [67]. This mechanism of interaction between soluble-Flt1 and VEGFA as well as soluble-Eng and TGFB1 (also present in community 2) had being long term related with PE pathogenesis [66]. The increment in VEGFA could also be associated with an increment in HSP90, also acting with SRC in the NOS3 expression and NO production. As we previously explained, this can also be accomplished through PI3K and the involvement of AKT1. The expression of HSP90 could be polemic because apparently it is related with the disease progression

and also placenta location [68]. However several authors had being found an increment in the placenta expression of HSP90 in PE [68–70] at mRNA and protein levels. Moreover, this increment can be a protective reaction that is stimulated by HIF1A [71] and consequently connected to disease progression as described in [68]. Actually, there are differences in HIF1A placenta expression in early-onset PE and late-onset PE [72] showing that both stages have different hypoxia compensatory mechanism. Interestingly the HIF1A gene is part of our prioritized list (ranking at 46) and PPI network. HIF1A is not part of any community but connected to several of them, especially with community 2 (HIF1A is not connected with communities 1 and 7) and it is also missing in the integrated metabolic models. There are not studies of HSP90 (or closely related HSP70) gene variations or promoter polymorphism in PE to know for sure if the protective roll could be compromised leading to early or late PE manifestation.

In the extended model (Model 3, in Additional file 5) the VEGF pathways is connected with several members of the HLA family through CD247 and PAK2 but also to another variety of pathways through ROCK2 and FGD3. Both connections can be related with apoptosis, trophoblastic affectation and endothelial cells organization [73, 74]. Even when PAK2 had being poorly studied in PE, we know that it is directly involved in gestational trophoblastic disease [45] and that endothelial cells PAK and/or CDC42 are directly involved with KDR and consequently essential for endothelial cells organization [74, 75].

The roll of angiogenesis in PE pathogenesis is clearly revealed in our theoretical analysis as well as scientific literature. However, we should discuss other aspects related with pathogenesis, specially the renin-angiotensin pathway and the roll of catechol-O-methyltransferase (COMT).

We can notice in our prioritization list that AGT and AGTR1 are the first two genes in our ranking (ACE and AGTR2 were also found in position 7 and 19 respectively) but interestingly only AGTR1 was found down-regulated in our microarray data. The down-regulation of AGTR1 from microarray data was only identified (considering our microarray data) in [6, 8], however, other authors [76, 77] found an increment in AGTR1 expression. In any case there are evidences that renin-angiotensin system is modified in hypoxic conditions [77, 78] and it is well connected with the HIF1A previously discussed. In our list of pathogenic genes, we should notice the AGT derive from 1) the AT₁-AA auto-antibody that interact with AGTR1 or 2) from some polymorphism in AGT that had being associated with increased risk of preeclampsia [54, 79, 80]. The origin of AT₁-AA in PE is quite unknown [50, 51, 81] but some authors shown that it is related to B-cells and it is connected with IL10 during pregnancy as well as with other cytokines (i.e. TNF) [82, 83]. Other authors indicates that an increment in CD4(+) T-cells and a decrement in T

Table 8 Compound list of metabolic species present in the expanded integrated metabolic network model

3-Oxopalmitoyl-CoA	C05259	Deoxyadenosine	C00559
Corticosterone	C02140	Deoxyinosine	C05512
Arachidonate	C00219	Tetrahydrofolate (THF)	C00101
3alpha,7alpha-Dihydroxy-5beta-cholestanoyl-CoA	C04644	Bromobenzene-3,4-oxide	C14839
11beta-Hydroxyandrost-4-ene-3,17-dione	C05284	Parathion (DNTP)	C06604
3alpha,7alpha-Dihydroxy-5beta-cholestanate;	C04554	Oxriptan	C00643
2-Methoxy-17beta-estradiol	C05302	Bilirubin	C00486
Triglyceride	C00422	Docosahexaenoic acid (DHA)	C06429
5(S)-HPETE	C05356	Nicotinamide mononucleotide (NMN)	C00455
16alpha-Hydroxyestrone	C05300	dAMP	C00360
Prostaglandin G2	C05956	Melatonin	C01598
L-Homocysteine	C00155	Nicotinamide	C00153
Glutathione (GSH)	C00051	Serotonin	C00780
Serine	C00065	N(omega)-Hydroxyarginine;	C05933

regulatory cells stimulates TNF, IL6, endothelin (EDN1), IL-17 and B-cells production of AT₁-AA [83–85]. Some of these genes are clearly involved in our networks and communities, especially EDN1 and IL6. We can't clearly state that renin-angiotensin pathway is not part of PE pathogenesis but our network based results reduce the importance of this pathway. These suggest that its finding in the top prioritized gene list is actually a consequence of the hypertension effect more than the PE pathogenesis.

The additional consideration of the compounds involved in the integrated metabolic network (Model 3) (Table 8) also leads us to interesting points. A deregulation in ammonia and urea cycles as well as in phospholipid and bile acid metabolism has being reported previously in metabolomics analysis [86, 87]. We know that steroid hormones are related with vascular endothelium, for instance, estradiol and progesterone/estradiol ratios are altered in placenta of PE women probably related with NO metabolism [88]. However, one of the most relevant results in this Table 9 and in the expanded integrated metabolic network (Model 3) is the presence of catechol-O-methyltransferase (COMT) and 2-methoxyestradiol. In our prioritized gene list (see Additional file 3) the COMT gene is ranked at position 47 and we know

that pregnant mouse with a deficiency in catechol-O-methyltransferase (COMT) (consequently no 2-methoxyestradiol) lead to PE phenotype [89]. This animal model was not considered in our initial pathogenic data analysis but it is clearly expressed in the integrated pathway and prioritization strategy. In the model 3 (Additional file 5), the COMT gene is quite far connected with VEGF, however, was recently showed that 2-methoxyestradiol has an anti-angiogenic effect connected to KDR and HIF1A probably through a different mechanism not involving sFlt-1 [90].

Our results confirm that consensus prioritization strategy lead us to genes with pathogenic involvement, at least in PE. Moreover, the introductions of network and enrichment analysis are capable to narrow the metabolic and gene space leading us toward reasonable conclusions in agreement with our scientific knowledge of the disease. However, the proposed strategies need to be further improved in several topics. For instance: a) the inclusion of prioritization algorithms based in learning strategies, b) the inclusion of other network processing methods to reduce the gene lost and 3) the differentiation between early and late-onset preeclampsia. Additionally, as previously stated, there are several genes relevant in our analysis with poor or almost no information in their PE involvement. Therefore further experimental analysis will needed to validate the participation of these genes in PE pathogenesis or clinical manifestation.

Table 9 Results of integrated metabolic pathways

Models	Number of initial genes	p-value
Model 1 (0 missing gene(s) are allowed)	98	< 0.005
Model 2 (1 missing gene(s) are allowed)	209	< 0.005
Model 3 (2 missing gene(s) are allowed)	231	< 0.005

Conclusions

From all the prioritization methods used in our work MetaRanker brings the better results. However, our results confirm that consensus strategy of several prioritization tools improve the detection and initial enrichment of pathogenic genes, at least in preeclampsia condition.

The combination of around the first percent of the prioritized genes and protein-protein interaction network followed by communality analysis brings the possibilities to reduce the gene space and actually group well known genes related with pathogenesis. In this analysis communities connected with VEGF-signaling pathway are highly enriched. This pathway is also enriched considering the microarray data. Actually the pathways weighting strategy together with network analysis agrees with the results obtained in microarray data.

The integrated metabolic pathway clearly indicates main routes involved in preeclampsia pathogenesis. Our result could support previous publications indicating that hypoxia and also angiotensin pathways are secondary manifestations and could be actually connected with disease progression or differentiation between early and late onset preeclampsia development. Our result point to VEGF, FLT1, KDR as relevant pathogenic genes, as well as those connected with NO metabolism. However, other genes like HSP90, PAK2, CD247 and others included in the first 1% of the prioritized list need to be further explored in preeclampsia pathogenesis through experimental approaches.

Additional files

Additional file 1: Identification of pathogenic genes. The file comprises the literature and several observations considered for the selection of our pathogenic gene list. (DOCX 86 kb)

Additional file 2: Microarrays consensus. The file comprises all information concerning the microarray data as well as the integration. (XLSX 141 kb)

Additional file 3: Prioritized genes. The file comprises our final prioritized genes as well as the consensus score. (XLSX 28 kb)

Additional file 4: Enrichment analysis. The file comprises all the enrichment analysis: gene ontology and metabolic pathways. (XLSX 292 kb)

Additional file 5: Integrated metabolic network. The file comprises the Integrated Metabolic Network corresponding with Model 3 of Table 8 as well as the list of all compounds contained in the metabolic network. (DOCX 1116 kb)

Abbreviations

AUAC: Under the accumulation curve; BEDROC: Boltzmann-enhanced discrimination of ROC; EF: Enrichment factor; PE: Preeclampsia; RIE: Robust initial enhancement (RIE); ROC: Receiver operating curve

Funding

This project was partially supported by Foundation for Science and Technology (FCT) and FEDER/COMPETE (Grants UID/QUI/00081/2013, POCI-01-0145-FEDER-006980, and NORTE-01-0145-FEDER-000028). The authors also thank the COST action CA15135 (Multi-Target Paradigm for Innovative Ligand Identification in the Drug Discovery Process, MuTaLig) for support. MC-M (Grant SFRH/BPD/90673/2012) was also supported by FCT and FEDER/COMPETE funds.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' contributions

ET, MC, AS and YPC were responsible for methodologies and algorithms development. Specifically, ET and MC worked in the consensus strategy and

metabolic network while YPC and AS worked in protein-protein interaction network. GB and ME were involved in data recollection from literature concerning to gene-disease confirmed associations as well as microarray data integration. FB, NC and CP were involved genes-disease validation analysis and results discussion. IR was involved in metabolic analysis and genetic results integration with possible interpretation in preeclampsia. All authors were involved in manuscript corrections. All authors had read and approved the manuscript for publication.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Facultad de Medicina, Universidad de Las Américas, Av. de los Granados E12-41y Colimes esq, EC170125 Quito, Ecuador. ²Department of Molecular and Cellular Pharmacology, Miller School of Medicine and Center for Computational Science, University of Miami, FL 33136 Miami, USA. ³Department of General Education, West Coast University—Miami Campus, Doral FL 33178, USA. ⁴Departamento de Ciencias Naturales, Universidad Técnica Particular de Loja, Calle París S/N, EC1101608, Loja, Ecuador. ⁵Escuela de Ciencias Físicas y Matemáticas, Universidad de Las Américas, Quito, Ecuador. ⁶CIQUP/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal. ⁷REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal. ⁸Centro de Investigaciones genética y genómica, Facultad de Ciencias de la Salud, Universidad Tecnológica Equinoccial, Quito, Ecuador. ⁹Faculty of Pharmacy, University of Porto, Porto, Portugal. ¹⁰UCIBIO@REQUIMTE, Caparica, Portugal.

Received: 8 February 2017 Accepted: 28 July 2017

Published online: 08 August 2017

References

- Chaiworapongsa T, Chaemsaitong P, Yeo L, Romero R. Pre-eclampsia part 1: current understanding of its pathophysiology. *Nat Rev Nephrol*. 2014;10:466–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25003615>
- Fisher SJ. Why is placentation abnormal in preeclampsia? *Am J Obstet Gynecol*. 2015;213:S115–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26428489>
- Zhou Y, Damsky CH, Fisher SJ. Preeclampsia is associated with failure of human cytotrophoblasts to mimic a vascular adhesion phenotype. One cause of defective endovascular invasion in this syndrome? *J Clin Invest*. 1997;99:2152–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9151787>
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*. 2011;39:D1005–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21097893>
- Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, et al. ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res*. 2011;39:D1002–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21071405>
- van Uiter M, Moerland PD, Enquobahrie DA, Laiuuri H, van der Post JAM, Ris-Stalpers C, et al. Meta-analysis of placental Transcriptome data identifies a novel molecular pathway related to preeclampsia. *PLoS One*. 2015;10:e0132468. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26171964>
- Yong HEJ, Melton PE, Johnson MP, Freed KA, Kalionis B, Murthi P, et al. Genome-wide transcriptome directed pathway analysis of maternal preeclampsia susceptibility genes. *PLoS One*. 2015;10:e0128230. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26010865>
- Leavey K, Bainbridge SA, Cox BJ. Large scale aggregate microarray analysis reveals three distinct molecular subclasses of human preeclampsia. *PLoS*

- One. 2015;10:e0116508. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25679511>
9. Rabaglino MB, Post Uiterweer ED, Jeyabalan A, Hogge WA, Conrad KP. Bioinformatics approach reveals evidence for impaired endometrial maturation before and during early pregnancy in women who developed preeclampsia. *Hypertension*. 2015;65:421–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25421975>
 10. Tejera E, Bernardes J, Rebelo I. Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia. *BMC Med Genet*. 2013;6:51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24219996>
 11. Tejera E, Bernardes J, Rebelo I. Preeclampsia: a bioinformatics approach through protein-protein interaction networks analysis. *BMC Syst Biol*. 2012;6:97. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22873350>
 12. Song Y, Liu J, Huang S, Zhang L. Analysis of differentially expressed genes in placental tissues of preeclampsia patients using microarray combined with the connectivity map database. *Placenta*. 2013;34:1190–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24125805>
 13. Vaiman D, Calicchio R, Miralles F. Landscape of transcriptional deregulations in the preeclamptic placenta. *PLoS One*. 2013;8:e65498. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23785430>
 14. Moslehi R, Mills JL, Signore C, Kumar A, Ambroggio X, Dzusev A. Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. *Sci Rep*. 2013;3:2407. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23989136>
 15. Börnigen D, Tranchevent L-C, Bonachela-Capdevila F, Devriendt K, De Moor B, De Causmaecker P, et al. An unbiased evaluation of gene prioritization tools. *Bioinformatics*. 2012;28:3081–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23047555>
 16. Tranchevent L-C, Capdevila FB, Nitsch D, De Moor B, De Causmaecker P, Moreau Y. A guide to web tools to prioritize candidate genes. *Brief Bioinform*. 2011;12:22–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21278374>
 17. Liekens AML, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J. BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. *Genome Biol*. 2011;12:R57. *BioMed Central*, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21696594>
 18. Hutz JE, Kraja AT, McLeod HL. Province MA. CANDID: a flexible method for prioritizing candidate genes for complex human traits. *Genet Epidemiol*. 2008;32:779–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18613097>
 19. Jourquin J, Duncan D, Shi Z, Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics*. 2012;13(Suppl 8):S20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23282288>
 20. Cheng D, Knox C, Young N, Stothard P, Damaraju S, Wishart DS. PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Res*. 2008;36:W399–405. Oxford University Press, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18487273>
 21. Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. *Mol Syst Biol*. 2008;4:189. European Molecular Biology Organization, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18463613>
 22. Guney E, Garcia-Garcia J, Oliva B. GUILDify: a web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms. *Bioinformatics*. 2014;30:1789–90. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24532728>
 23. Piñero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)*. 2015;2015:bav028. Oxford University Press, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25877637>
 24. Yu W, Wulf A, Liu T, Khoury MJ, Gwinn M, Rebbeck T, et al. Gene prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases. *BMC Bioinformatics*. 2008;9:528. *BioMed Central*, Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-528>
 25. Fontaine J-F, Priller F, Barbosa-Silva A, Andrade-Navarro MA. Génie: literature-based gene prioritization at multi genomic scale. *Nucleic Acids Res*. 2011;39:W455–61. Oxford University Press, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21609954>
 26. Yue P, Melamud E, Moulit J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*. 2006;7:166. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16551372>
 27. Seelow D, Schwarz JM, Schuelke M. GeneDistiller—distilling candidate genes from linkage intervals. *PLoS One*. 2008;3:e3874. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19057649>
 28. Pers TH, Dworzynski P, Thomas CE, Lage K, Brunak S. MetaRanker 2.0: a web server for prioritization of genetic variation data. *Nucleic Acids Res*. 2013;41:W104–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23703204>
 29. Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent advances and emerging applications in text and data Mining for Biomedical Discovery. *Brief Bioinform*. 2016;17:33–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26420781>
 30. Helguera AM, Perez-Castillo Y, Cordeiro MN DS, Tejera E, Paz-Y-Miño C, Sánchez-Rodríguez A, et al. Ligand-based virtual screening using tailored ensembles: a prioritization tool for dual A2AAdenosine receptor antagonists / monoamine Oxidase B inhibitors. *Curr Pharm Des*. 2016;22:3082–96. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26932160>
 31. Perez-Castillo Y, Helguera AM, Cordeiro MNDS, Tejera E, Paz-Y-Miño C, Sánchez-Rodríguez A, et al. Fusing docking scoring functions improves the virtual screening performance for discovering Parkinson's disease dual target Ligands. *Curr Neuropharmacol*. 2017 [cited 2017 Mar 29]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28067172>
 32. Truchon J-F, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. *J Chem Inf Model*. 2009;9:474–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17288412>
 33. Cruz-Monteaugudo M, Borges F, Paz-y-Miño C, Cordeiro MNDS, Rebelo I, Perez-Castillo Y, et al. Efficient and biologically relevant consensus strategy for Parkinson's disease gene prioritization. *BMC Med Genet*. 2016;9:12. *BioMed Central*, Available from: <http://www.biomedcentral.com/1755-8794/9/12>
 34. Mackey MD, Melville JL. Better than random? The chemotype enrichment problem. *J Chem Inf Model*. 2009;9:1154–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19397275>
 35. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19131956>
 36. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19033363>
 37. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21789182>
 38. Antonov AV, Schmidt EE, 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 Res*. 2010;38:W78–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20519200>
 39. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25352553>
 40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14597658>
 41. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*. 2005;435:814–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15944704>
 42. Walsh CJ, Hu P, Batt J, Dos SCC. Microarray meta-analysis and cross-platform normalization: integrative genomics for robust biomarker discovery. *Microarrays (Basel, Switzerland)*. 2015;4:389–406. Multidisciplinary Digital Publishing Institute (MDPI), Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27600230>
 43. Cox B. Bioinformatic approach to the genetics of preeclampsia. *Obstet Gynecol*. 2014;124:633. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25162267>
 44. Jia R, Li J, Rui C, Ji H, Ding H, Lu Y, et al. Comparative proteomic profile of the human umbilical cord blood Exosomes between normal and preeclampsia pregnancies with high-resolution mass spectrometry. *Cell Physiol Biochem*. 2015;36:2299–306. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26279434>
 45. Tejera E, Bernardes J, Rebelo I. Preeclampsia: a bioinformatics approach through protein-protein interaction networks analysis. *BMC Syst Biol*. 2012;2012:97.

46. Khangura RK, Khangura CK, Desai A, Goyert G, Sangha R. Metastatic colorectal cancer resembling severe preeclampsia in pregnancy. *Case Rep Obstet Gynecol.* 2015;2015:487824. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26770850>
47. Romero R, Grivel J-C, Tarca AL, Chaemsaitong P, Xu Z, Fitzgerald W, et al. Evidence of perturbations of the cytokine network in preterm labor. *Am J Obstet Gynecol.* 2015;213:836.e1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26232508>
48. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16632515>
49. Iriyama T, Wang W, Parchim NF, Song A, Blackwell SC, Sibai BM, et al. Hypoxia-independent upregulation of placental hypoxia inducible factor-1 α gene expression contributes to the pathogenesis of preeclampsia. *Hypertension.* 2015;65:1307–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25847948>
50. Xia Y, Kellems RE. Angiotensin receptor agonistic autoantibodies and hypertension: preeclampsia and beyond. *Circ Res.* 2013;113:78–87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23788505>
51. Parrish MR, Murphy SR, Rutland S, Wallace K, Wenzel K, Wallukat G, et al. The effect of immune factors, tumor necrosis factor- α , and agonistic autoantibodies to the angiotensin II type I receptor on soluble fms-like tyrosine-1 and soluble endoglin production in response to hypertension during pregnancy. *Am J Hypertens.* 2010;23:911–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20431529>
52. Maynard SE, Min J-Y, Merchan J, Lim K-H, Li J, Mondal S, et al. Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J Clin Invest.* 2003;111:649–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12618519>
53. Venkatesha S, Toporsian M, Lam C, Hanai J, Mammoto T, Kim YM, et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. *Nat Med.* 2006;12:642–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16751767>
54. Staines-Urias E, Paez MC, Doyle P, Dudbridge F, Serrano NC, Ioannidis JPA, et al. Genetic association studies in pre-eclampsia: systematic meta-analyses and field synopsis. *Int J Epidemiol.* 2012;41:1764–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23132613>
55. Li X, Shen L, Tan H. Polymorphisms and plasma level of transforming growth factor-Beta 1 and risk for preeclampsia: a systematic review. *PLoS One.* 2014;9:e97230. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24823830>
56. Macintire K, Tuohy L, Ye L, Palmer K, Gantier M, Tong S, et al. PAPP2 is increased in severe early onset preeclampsia and upregulated with hypoxia. *Reprod Fertil Dev.* 2014;26:351–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23484525>
57. Wagner PK, Otomo A, Christians JK. Regulation of pregnancy-associated plasma protein A2 (PAPP2) in a human placental trophoblast cell line (BeWo). *Reprod Biol Endocrinol.* 2011;9:48. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21496272>
58. Fong FM, Sahemey MK, Hamed G, Eytayo R, Yates D, Kuan V, et al. Maternal genotype and severe preeclampsia: a HuGE review. *Am J Epidemiol.* 2014;180:335–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25028703>
59. Nezi M, Mastorakos G, Mouslech Z. Corticotropin releasing hormone and the immune/inflammatory response [internet]. *Endotext.* 2000. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25905246>
60. Song J, Li Y, An RF. Identification of early-onset preeclampsia-related genes and MicroRNAs by bioinformatics approaches. *Reprod Sci.* 2015;22:954–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25717061>
61. Noris M, Perico N, Remuzzi G. Mechanisms of disease: pre-eclampsia. *Nat Clin Pract Nephrol.* 2005;1:98–114. Nature Publishing Group, Available from: <http://www.nature.com/doi/10.1038/ncpneph0035>
62. Dimmeler S, Fleming I, Fisslthaler B, Hermann C, Busse R, Zeiher AM. Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation. *Nature.* 1999;399:601–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10376603>
63. Cindrova-Davies T, Sanders DA, Burton GJ, Charnock-Jones DS. Soluble FLT1 sensitizes endothelial cells to inflammatory cytokines by antagonizing VEGF receptor-mediated signalling. *Cardiovasc Res.* 2011;89:671–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21139021>
64. Nagai A, Sado T, Naruse K, Noguchi T, Haruta S, Yoshida S, et al. Antiangiogenic-induced hypertension: the molecular basis of signaling network. *Gynecol Obstet Investig.* 2012;73:89–98. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22222493>
65. Chappell JC, Taylor SM, Ferrara N, Bautch VL. Local guidance of emerging vessel sprouts requires soluble Flt-1. *Dev Cell.* 2009;17:377–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19758562>
66. Powe CE, Levine RJ, Karumanchi SA. Preeclampsia, a disease of the maternal endothelium: the role of antiangiogenic factors and implications for later cardiovascular disease. *Circulation.* 2011;123:2856–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21690502>
67. Sundrani DP, Reddy US, Joshi AA, Mehendale SS, Chavan-Gautam PM, Hardikar AA, et al. Differential placental methylation and expression of VEGF, FLT-1 and KDR genes in human term and preterm preeclampsia. *Clin Epigenetics. BioMed Central.* 2013;5:6. Available from: <http://clinalepigeneticsjournal.biomedcentral.com/articles/10.1186/1868-7083-5-6>
68. Hromadnikova I, Dvorakova L, Kotlabova K, Kestlerova A, Hymanova L, Novotna V, et al. Assessment of placental and maternal stress responses in patients with pregnancy related complications via monitoring of heat shock protein mRNA levels. *Mol Biol Rep.* 2015;42:625–37. Available from: <http://link.springer.com/10.1007/s11033-014-3808-z>
69. Shu C, Liu Z, Cui L, Wei C, Wang S, Tang JJ, et al. Protein profiling of preeclampsia placental tissues. Buratti E, editor. *PLoS One.* 2014;9:e112890. Public Library of Science, Available from: <http://dx.plos.org/10.1371/journal.pone.0112890>
70. Padmini E, Venkatraman U, Srinivasan L. Mechanism of JNK signal regulation by placental HSP70 and HSP90 in endothelial cell during preeclampsia. *Toxicol Mech Methods.* 2012;22:367–74. Available from: <http://www.tandfonline.com/doi/full/10.3109/15376516.2012.673091>
71. Padmini E, Uthra V, Lavanya S. Effect of HSP70 and 90 in modulation of JNK, ERK expression in Preeclamptic placental endothelial cell. *Cell Biochem Biophys.* 2012;64:187–95. Available from: <http://link.springer.com/10.1007/s12013-012-9371-0>
72. Khodzhaeva ZS, Kogan YA, Shmakov RG, Klimchenko NI, Akatyeva AS, Vavina OV, et al. Clinical and pathogenetic features of early- and late-onset pre-eclampsia. *J Matern Neonatal Med.* 2015;2015:1–7. Available from: <http://www.tandfonline.com/doi/full/10.3109/14767058.2015.1111332>
73. Siu MKY, Yeung MCW, Zhang H, Kong DSH, Ho JWK, Ngan HYS, et al. p21-activated kinase-1 promotes aggressive phenotype, cell proliferation, and invasion in gestational trophoblastic disease. *Am J Pathol.* 2010;176:3015–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20413688>
74. Barry DM, Xu K, Meadows SM, Zheng Y, Norden PR, Davis GE, et al. Cdc42 is required for cytoskeletal support of endothelial cell adhesion during blood vessel formation in mice. *Development.* 2015;142:3058–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26253403>
75. Dubrac A, Genet G, Ola R, Zhang F, Pibouin-Fragner L, Han J, et al. Targeting NCK-mediated endothelial cell front-rear polarity inhibits Neovascularization. *CLINICAL PERSPECTIVE. Circulation.* 2016;133:409–21. Available from: <http://circ.ahajournals.org/lookup/doi/10.1161/CIRCULATIONAHA.115.017537>
76. Mistry HD, Kurlak LO, Broughton Pipkin F. The placental renin-angiotensin system and oxidative stress in pre-eclampsia. *Placenta.* 2013;34:182–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23246097>
77. Kurlak LO, Williams PJ, Bulmer JN, Broughton Pipkin F, Mistry HD. Placental expression of adenosine A2A receptor and hypoxia inducible factor-1 α in early pregnancy, term and pre-eclamptic pregnancies: interactions with placental renin-angiotensin system. *Placenta.* 2015;36:611–3. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0143400415008103>
78. Kurlak LO, Mistry HD, Cindrova-Davies T, Burton GJ, Broughton Pipkin F. Human placental renin-angiotensin system in normotensive and pre-eclamptic pregnancies at high altitude and after acute hypoxia-reoxygenation insult. *J Physiol.* 2016;594:1327–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26574162>
79. Ni S, Zhang Y, Deng Y, Gong Y, Huang J, Bai Y, et al. AGT M235T polymorphism contributes to risk of preeclampsia: evidence from a meta-analysis. *J Renin-Angiotensin-Aldosterone Syst.* 2012;13:379–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22513276>
80. Zhao L, Dewan AT, Bracken MB. Association of maternal AGTR1 polymorphisms and preeclampsia: a systematic review and meta-analysis. *J Matern Fetal Neonatal Med.* 2012;25:2676–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22758920>
81. Dechend R, Gratz P, Wallukat G, Shagdarsuren E, Plehm R, Bräsen J-H, et al. Agonistic autoantibodies to the AT1 receptor in a transgenic rat model of preeclampsia. *Hypertension.* 2005;45:742–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15699466>

82. Fettke F, Schumacher A, Costa S-D, Zenclussen AC. B cells: the old new players in reproductive immunology. *Front Immunol.* 2014;5:285. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25002862>.
83. Spradley FT, Palei AC, Granger JP. Immune mechanisms linking obesity and preeclampsia. *Biomol Ther.* 2015;5:3142–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26569331>.
84. Harmon AC, Cornelius DC, Amaral LM, Faulkner JL, Cunningham MW, Wallace K, et al. The role of inflammation in the pathology of preeclampsia. *Clin Sci (Lond).* 2016;130:409–19. Portland Press Limited, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26846579>.
85. Dhillon P, Wallace K, Herse F, Scott J, Wallukat G, Heath J, et al. IL-17-mediated oxidative stress is an important stimulator of AT1-AA and hypertension during pregnancy. *Am J Phys Regul Integr Comp Phys.* 2012;303:R353–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22718806>.
86. Austdal M, Thomsen LCV, Tangerås LH, Skei B, Mathew S, Bjørge L, et al. Metabolic profiles of placenta in preeclampsia using HR-MAS MRS metabolomics. *Placenta.* 2015;36:1455–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26582504>.
87. Bahado-Singh RO, Syngelaki A, Akolekar R, Mandal R, Bjondahl TC, Han B, et al. Validation of metabolomic models for prediction of early-onset preeclampsia. *Am J Obstet Gynecol.* 2015;213:530.e1–530.e10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26116099>.
88. Zheng J-J, Wang H-O, Huang M, Zheng F-Y. Assessment of ADMA, estradiol, and progesterone in severe preeclampsia. *Clin Exp Hypertens.* 2016;38:347–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27152507>.
89. Kanasaki K, Palmsten K, Sugimoto H, Ahmad S, Hamano Y, Xie L, et al. Deficiency in catechol-O-methyltransferase and 2-methoxyoestradiol is associated with pre-eclampsia. *Nature.* 2008;453:1117–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18469803>.
90. Lee DK, Nevo O. 2-Methoxyestradiol regulates VEGFR-2 and sFlt-1 expression in human placenta. *Placenta.* 2015;36:125–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25499009>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

