Computational characterization and identification of human polycystic ovary syndrome genes

Xing-Zhong Zhang¹, Yan-Li Pang³, Xian Wang¹ & Yan-Hui Li²

Human polycystic ovary syndrome (PCOS) is a highly heritable disease regulated by genetic and environmental factors. Identifying PCOS genes is time consuming and costly in wet-lab. Developing an algorithm to predict PCOS candidates will be helpful. In this study, for the first time, we systematically analyzed properties of human PCOS genes. Compared with genes not yet known to be involved in PCOS regulation, known PCOS genes display distinguishing characteristics: (i) they tend to be located at network center; (ii) they tend to interact with each other; (iii) they tend to enrich in certain biological processes. Based on these features, we developed a machine-learning algorithm to predict new PCOS genes. 233 PCOS candidates were predicted with a posterior probability >0.9. Evidence supporting 7 of the top 10 predictions has been found.

Polycystic ovary syndrome (PCOS) is a highly complex disorder that affects 6–10% of women of reproductive age¹. It is a major cause of anovulatory infertility and increases the risk for insulin resistance, obesity, cardiovascular disease and psychosocial disorders²,³. Studies have shown that PCOS is regulated by the subtle interaction of genes and environmental factors⁴–⁶.

To identify PCOS genes, reverse genetics like microarray studies have profiled whole-genome gene expression in a number of PCOS tissues, including ovary⁷,⁸ and adipose⁹. Genome-wide association study (GWAS) is used to identify regions of the genome that harbor variants associated with disease risk or quantitative traits¹⁰–¹². For computational methods, a group once reconstructed transcription factor-microRNA synergistic regulatory network, and they considered the nodes with highest degree as PCOS candidate genes¹³. Another group constructed a protein-protein interaction (PPI) subnetwork and selected the top hubs (both high degree and betweenness) as PCOS candidates¹⁴. However, both works lack rigorous statistics to evaluate the accuracy of the prediction. To our knowledge, no efficient algorithm has been developed to predict PCOS genes. In fact, bioinformatics algorithms have been successfully developed to infer candidate genes in other fields¹⁵–¹⁸, and these could be introduced to PCOS research.

In this work, we developed a method to identify distinguishing properties of PCOS genes and subsequently used them to predict new candidates. We firstly systematically compared the computational characteristics of two groups of genes: known PCOS genes versus the remaining genes in the genome (called non-PCOS genes hereafter). We examined each set of the genes in network topological features and functional annotations. Then, we singled out the features with significant difference between PCOS and non-PCOS genes by Kolmogorov–Smirnov (KS) test. We employed support vector machine (SVM) with linear function as the classifier. Finally, with a posterior probability >0.9, 233 new PCOS genes were predicted. Literature supporting 7 of the top 10 predictions has been found.

Results

PCOS genes tend to have higher degrees. For a protein, its degree is defined as the number of direct interaction genes. According to network theory, a protein with more direct interaction neighbors (higher degree) might be more important to the network¹⁹. Based on PPI network downloaded from OPHID²⁰, we counted the number of direct interaction neighbors for each gene, and found that PCOS genes tend to have higher degrees than non-PCOS genes. The average degrees for PCOS genes and non-PCOS genes are 41.81 and 22.48, respectively (see Table 1). The cumulative frequency distribution curves of degrees for PCOS genes shift to the right compared with that of non-PCOS genes (Fig. 1A). There is significant difference between them, with \( P = 4.2E-13 \) by KS test.

¹Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Peking University, Beijing, China. ²Institute of Cardiovascular Sciences, Peking University, Beijing, China. ³Department of Obstetrics and Gynecology, Center for Reproductive Medicine, Peking University Third Hospital, Beijing, China. Correspondence and requests for materials should be addressed to X.W. (email: xwang@bjmu.edu.cn) or Y.-H.L. (email: liyanhui@bjmu.edu.cn)
To validate the results, we also analyzed degrees of PCOSDB genes or PCOSKB genes separately. The average degree of PCOSDB genes is 50.54 while that from PCOSKB equals to 38.60, which are all significantly higher than that of non-PCOS genes. These results were listed in Table 1.

**PCOS genes tend to enrich at global network center.** Genes with high degrees might locate at globally or locally central position, while only those located globally are more likely to be evolutionarily conserved21. To distinguish the different locations, we calculated K-core for each gene. The K-core gradually displays the backbone of a network by iteratively deleting genes with a degree lower than K, until the remaining genes in the subnetwork have a degree higher than K.

Betweenness centrality measures the number of times that a gene is on the shortest path between two other genes. Therefore, a gene with a high betweenness could be considered as a bottleneck node in the network22. The results showed that PCOS genes had significantly higher betweenness along the shortest path between two genes than that of non-PCOS genes, with the average betweennesses are 34,463 and 17,278 (P = 1.4E-17 by KS test; Table 1 and Fig. 1C). And as shown in Table 1, PCOS genes have similar average betweenness from the other two datasets.

| Dataset  | Class  | Size  | Degree  | K-Core  | Betweenness | 1st PCOS Ratio | 2nd PCOS Ratio |
|----------|--------|-------|---------|---------|-------------|--------------|--------------|
| Total    | PCOS   | 306   | 41.81   | 16.78   | 34463       | 0.11         | 0.04         |
|          | Non-PCOS | 16676 | 22.48   | 11.66   | 17278       | 0.04         | 0.03         |
|          |        | P value | 4.2E-13 | 1.7E-10 | 1.4E-17     | 3.0E-48      | 6.0E-20      |
| PCOSDB   | PCOS   | 185   | 50.54   | 18.78   | 40177       | 0.10         | 0.03         |
|          | Non-PCOS | 16676 | 22.48   | 11.66   | 17278       | 0.02         | 0.02         |
|          |        | P value | 1.5E-12 | 5.2E-10 | 2.0E-12     | 2.0E-36      | 1.3E-09      |
| PCOSKB   | PCOS   | 226   | 38.60   | 15.43   | 33614       | 0.09         | 0.04         |
|          | Non-PCOS | 16676 | 22.48   | 11.66   | 17278       | 0.03         | 0.02         |
|          |        | P value | 2.6E-07 | 9.1E-07 | 5.3E-12     | 1.5E-32      | 1.1E-20      |

Table 1. Network Characteristics of PCOS Genes. “Total” indicates all the PCOS genes covered by either PCOSDB or PCOSKB. “Non-PCOS” indicates the remaining genes. The degree of a gene is defined as the number of its direct interaction genes. A K-core of a network can be obtained by recursively deleting genes with a degree lower than K, until the remaining genes in subnetwork have a degree higher than K. Betweenness centrality counts the number of times that a gene is on the shortest path between two other genes. 1st PCOS ratio is defined as the ratio of the number of PCOS genes that it directly interacts to its degree. 2nd PCOS ratio is defined as the ratio of the number of PCOS genes that belong to 2nd interaction genes to its number of 2nd interaction genes. The P values were calculated by KS test. PCOS genes represent polycystic ovary syndrome.

**GO functional enrichment.** As reported, genes associated with the same disease are often functionally related18,23. To examine whether PCOS genes tend to take part in some biological processes, a log-odds scores was computed for each GO term to compare the frequency at which PCOS genes and non-PCOS genes were annotated to it. The distributions of log-odds scores have a significant difference between PCOS genes and non-PCOS genes (P = 6.0E-20 by KS test). There results could be found in Table 1 and Fig. 1E.

No ovulation is a major diagnostic criterion for PCOS. As shown in Supplementary Table S1, “GO:0022602 ovulation cycle process” is enriched with PCOS genes. Steroid hormone plays an important role in ovarian development and ovulation process. Consistently, “GO:0042446 hormone biosynthetic process” is enriched. Meanwhile, “GO:0045940 positive regulation of steroid metabolic process” is significantly enriched with a log-odds score of 3.57, because steroid is a precursor for steroid hormone. These results indicate dysregulation of steroid hormone might be one major cause of PCOS. PCOS is a complex metabolic disease, and insulin resistance is another etiology. Thus, GO terms associated with regulation of plasma glucose are enriched, such as “GO:0048009 insulin-like growth factor receptor signaling pathway” and “GO:0010828 positive regulation of glucose transport”.

**Figure 1.** a) Network of PCOS genes. “Total” indicates all the PCOS genes covered by either PCOSDB or PCOSKB. “Non-PCOS” indicates the remaining genes. The degree of a gene is defined as the number of its direct interaction genes. A K-core of a network can be obtained by recursively deleting genes with a degree lower than K, until the remaining genes in subnetwork have a degree higher than K. Betweenness centrality counts the number of times that a gene is on the shortest path between two other genes. 1st PCOS ratio is defined as the ratio of the number of PCOS genes that it directly interacts to its degree. 2nd PCOS ratio is defined as the ratio of the number of PCOS genes that belong to 2nd interaction genes to its number of 2nd interaction genes. The P values were calculated by KS test. PCOS genes represent polycystic ovary syndrome.

**Table 1.** Network Characteristics of PCOS Genes. “Total” indicates all the PCOS genes covered by either PCOSDB or PCOSKB. “Non-PCOS” indicates the remaining genes. The degree of a gene is defined as the number of its direct interaction genes. A K-core of a network can be obtained by recursively deleting genes with a degree lower than K, until the remaining genes in subnetwork have a degree higher than K. Betweenness centrality counts the number of times that a gene is on the shortest path between two other genes. 1st PCOS ratio is defined as the ratio of the number of PCOS genes that it directly interacts to its degree. 2nd PCOS ratio is defined as the ratio of the number of PCOS genes that belong to 2nd interaction genes to its number of 2nd interaction genes. The P values were calculated by KS test. PCOS genes represent polycystic ovary syndrome.
The performance of the classifier. To model all above features, we tested different classifiers, K-nearest neighbor (KNN), decision tree and SVM with different kernel functions. As described in the Materials and methods, 306 PCOS genes downloaded from PCOSDB and PCOSKB were used as positive samples, and 306 negative genes were randomly sampled from the non-PCOS genes. Since random sampling might introduce bias, we sampled 1001 negative datasets and combined each negative dataset with the positive dataset to train the classifier. The median value of 1001 results was used to evaluate the performance of different classifiers. As shown in the Table 2, we found SVM (liner) achieved the best performance, with precision $= 0.81$, recall $= 0.71$, $F_1 = 0.75$, and $AUC = 0.80$. Thus, it was chosen as the final classifier and used for real application. To show the variances introduced by sampling, the boxplots of the 1001 training results (precisions, recalls, $F_1$s, and AUCs) with SVM (liner) are given in Supplementary Figure S1. The ROC curves of SVM (liner) could be found in Fig. 2.

Furthermore, we also tested the classifier with independent datasets. Firstly, we used the 226 PCOSKB genes as positive samples and took the 185 PCOSDB genes as test samples. We randomly selected 226 genes from non-PCOS ones as negative samples to train the classifier. After repeating 1001 times, we employed the model with median AUC value to predict the PCOSDB genes. Of the 79 PCOSDB genes (excluding the ones in the PCOSKB), 53, 32 and 15 genes were recalled with a posterior probability higher than 0.5, 0.8 and 0.9, respectively. This also showed the algorithm is helpful.

Real application. To predict new PCOS candidate genes, we took the 306 PCOS genes from PCOSDB and PCOSKB as positive samples. And the genes in the dataset that got median AUC value of the 1001 randomly selected datasets were taken as negative samples. After training the classifier, we found that 13,681 unknown genes could be predicted by the algorithm. With a posterior probability higher than 0.5, 0.8 and 0.9, respectively. This also showed the algorithm is helpful.

Figure 1. Cumulative frequency distributions of network features of PCOS genes and non-PCOS genes. The PCOS genes tend to have higher degree (A), K-core (B), betweenness (C), 1st PCOS ratio (D), and 2nd PCOS ratio (E) than that of non-PCOS genes. The cumulative frequency of different features is 100% for PCOS genes and non-PCOS genes. PCOS represents polycystic ovary syndrome.
SMAD3 is predicted as PCOS gene with a posterior probability = 0.9979. Allele rs11031006-G in SMAD3 was reported to be associated with lower PCOS risk.

Discussion

In this work, we systematically investigated properties of PCOS genes and then developed an algorithm to predict new PCOS genes by integrating network characteristics and GO functional characteristics. Different from GWAS and other genetic methods, this work opens a new avenue to infer PCOS candidates.

Previously, two methods have been reported to infer PCOS genes. One used degree as feature and the other used both degree and betweenness as features. In this work, besides degree and betweenness, we considered more network topological features like K-core, 1st PCOS ratio and 2nd PCOS ratio. And we integrated GO functional annotations to the algorithm. More important, our method is a supervised machine-learning algorithm, with rigorous statistics to evaluate the performance. And each predicted gene has a probability to evaluate the reliability of the prediction.

According to PCOSDB and PCOSKB, both IGF1 and IGF2 are PCOS genes. And in the PPI network, we found that most of their direct interaction neighbours are also PCOS genes. In addition, gene set enrichment analysis showed that IGF receptor signalling pathway (GO:0048009) is statistically enriched by PCOS genes, in which 18 of the 36 annotated genes are PCOS genes. And 12 of the rest 18 genes were predicted as PCOS candidates by our algorithm. These results are consistent with recent researches that IGF signalling pathways might play an important role in PCOS regulation.

It is known that PCOS is a highly heritable (70%) disease. However, to date, only one gene named PCOS1 has been collected to online Mendelian inheritance in man database. The PCOS genes analyzed in this work are downloaded from PCOSDB or PCOSKB. They are in fact PCOS-causing genes or PCOS-associated genes, since the causal relationships might need further confirmation by physiological studies. Here, we called them PCOS genes on one hand for short, on the other hand to highlight the importance of genetic background.

In this work, we defined 306 PCOS genes as positive samples and sampled 306 negative samples from the rest genes. We trained SVM model and evaluated classification performance with an equal number of positives and negatives, which has been widely adopted in previous studies. However, as mentioned by Myers et al., we should carefully interpret the results based on this method, because it is achieved under the assumption that the number of positives to the number of negatives equals to 1:1 in real application domain. Notably, it would also be improper if all remaining genes were selected as negatives. Because there might be not-yet-identified PCOS genes in negative samples, which might seriously underestimate the classifier.

Notably, current PPI data is far from perfect. They usually contain a number of false positive interactions and even more false negatives. Thus, some limitations are inevitable. For example, the degree of a protein might be related to the number of researches on it. And K-core, 1st and 2nd PCOS ratios might be indirectly related to such research bias. We think, with the improvement of PPI data quality, these problems could be solved and our approach could be more effective.
Materials and Methods

Data source. The PPI data were downloaded from the Online Predicted Human Interaction Database (OPHID; http://ophid.utoronto.ca/ophidv2.204/)20. After deleting self-interactions and redundant interactions, the final PPI network covers a total of 16,982 proteins and 193,949 edges. Two lists of PCOS genes were downloaded from the Polycystic Ovary Syndrome Database (PCOSDB; http://www.pcosdb.net/)32 and the KnowledgeBase on Polycystic Ovary Syndrome (PCOSKB; http://pcoskb.bicnirrh.res.in/)33, with 208 and 241 genes, respectively. 185 of the 208 PCOSDB genes, and 226 of the 241 PCOSKB genes were covered by the OPHID network. We combined the PCOSDB genes and the PCOSKB genes and got 306 PCOS genes in total. The functional annotations of gene products were obtained from gene ontology (GO) http://www.geneontology.org34. The source codes could be downloaded from Github: https://github.com/Heyuanshan/PCOS-genes-prediction.git.

Network topological features. The network features analyzed in this work, i.e., degree, K-core, betweenness and PCOS ratios (1st and 2nd), are defined in Table 4. They were computed by an R package, igraph35.

Log-odds score. We defined the log-odds score to describe the relative frequency with which a GO biological process was used to annotate PCOS or non-PCOS genes. The formula for calculation is as follows:

$$\text{Log-odds score} = \log \left( \frac{(m + a)/(n + a)}{m_0/n_0} \right)$$

$m_0$ is the number of PCOS genes; $n_0$ is the total number of genes in human genome; $m$ is the number of PCOS genes annotated to a GO term; and $n$ is the total number of human genes annotated to the GO term. $a (a = 1)$ is a correction factor. To avoid bias, we only used the GO terms annotated with more than 5 genes ($n \geq 5$). If a gene annotated to a GO term with a high log-odds score, then the gene is more likely a PCOS gene. If a gene is annotated to several GO terms, the log-odds scores of these GO terms were summed to reflect its total associations to PCOS.

Kolmogorov–Smirnov test. The Kolmogorov-Smirnov test is a useful nonparametric statistical method for comparing two samples through quantifying a distance between the empirical distribution functions of them. In this work, we used two sample KS test to compare the network features and functional annotations between PCOS genes and non-PCOS genes.

| Symbol | Name                                      | Posterior Probability |
|--------|-------------------------------------------|-----------------------|
| CTNNB1 | catenin beta 1                            | 0.99932               |
| THBS1  | thrombospondin 1                          | 0.99864               |
| IFNG   | interferon gamma                          | 0.99794               |
| SMAD3  | SMAD family member 3                      | 0.99736               |
| WNT5A  | Wnt family member 5 A                     | 0.99694               |
| EGFR   | epidermal growth factor receptor          | 0.9964                |
| HIF1A  | hypoxia inducible factor 1 subunit alpha  | 0.99623               |
| SRC    | SRC proto-oncogene, non-receptor tyrosine kinase | 0.99614            |
| ENG    | endoglin                                  | 0.99536               |
| NOG    | noggin                                    | 0.99505               |
| SIRT1  | sirtuin 1                                 | 0.99498               |
| PTEN   | phosphatase and tensin homolog            | 0.99429               |
| SHH    | sonic hedgehog                            | 0.9936                |
| CAV1   | caveolin 1                                | 0.9934                |
| SMAD4  | SMAD family member 4                      | 0.99129               |
| GREM1  | gremlin 1, DAN family BMP antagonist      | 0.9893                |
| BMP10  | bone morphogenetic protein 10             | 0.9886                |
| GDF5   | growth differentiation factor 5           | 0.98854               |
| FGA    | fibrinogen alpha chain                    | 0.98846               |
| GATA3  | GATA binding protein 3                    | 0.98752               |
| TGFBR3 | transforming growth factor beta receptor 3 | 0.98745               |
| JAK2   | Janus kinase 2                            | 0.98715               |
| LYN    | LYN proto-oncogene, Src family tyrosine kinase | 0.98662           |
| NOTCH1 | notch 1                                   | 0.98624               |
| LGA159 | galectin 9                                | 0.98621               |

Table 3. Top 25 Predicted PCOS Genes. Posterior probability is given by SVM to evaluate the reliability of the prediction. SVM represents support vector machine.
Table 4. Formal Representation of Graph Measures. Functions are the definitions of the topological features. Descriptions give explanations for symbols in the definitions.

| Name         | Function | Descriptions |
|--------------|----------|--------------|
| Degree       | K_i      | the number of direct interaction partners of node i |
| Degree-2     | K_i      | the number of 2-step interaction partners of node i |
| 1st PCOS ratio | K_{i,j}^{2} / K_i | K_{j}^{2} is the number of direct interactions between node i and proteins encoded by PCOS genes |
| 2nd PCOS ratio | K_{i,j}^{2} / K_i | K_{i,j}^{2} is the number of 2-step interaction between node i and proteins encoded by PCOS genes |
| Betweenness  | σ(j, k)   | σ(j, k) is the total number of shortest connections between nodes j and k, where each shortest connection has to pass node i, and σ(j, k) is the total number of shortest connections between j and k. The set V of nodes represents all proteins in the network. |
| K-core       | K        | A K-core of a graph can be obtained by recursively removing all nodes with a degree less than K, until all nodes in the remaining graph have a degree at least K. |

Classifiers. We tested different classifiers to predict PCOS genes: K-nearest neighbor (KNN), decision tree and SVM with different kernel functions. KNN and decision tree were employed from MATLAB, and SVM were employed from LIBSVM3.22. As shown in the Results, SVM with linear kernel achieved the best performance. The parameter c was optimized and set at 1. For each predicted gene, LIBSVM gives a posterior probability to evaluate its reliability. If a gene gets a larger posterior probability, then it is more likely a PCOS gene.

Positive and negative samples. The 306 PCOS genes obtained from PCOSDB and PCOSKB were used as positive samples. We randomly selected 306 genes from the rest of the human genome as the negative samples. This method has frequently been used to predict disease genes. To avoid sampling bias, we sampled 1001 times of the negative datasets of 306 genes, and combined each negative dataset with the positive dataset to train the classifier.

Classifier evaluation. As in previous works, we used 5-fold cross-validation to evaluate the classifier, in which 20 percent of the whole data were left out as the test data and the remaining (80 percent) as the training data. Precision, recall, F1 score, and area under curve (AUC) were used as the measures to evaluate the classifier. For each test dataset, we counted the numbers of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP). The formulas for calculating precision, recall, and F1 score were as following:

\[
\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Because we sampled 1001 negative datasets, and combined each negative dataset with the positive dataset to train the classifier, we got 1001 training results. We used the median of the 1001 values of precisions, recalls, F1s, and AUCs as the final results.

References
1. McCartney, C. R. & Marshall, J. C. Clinical Practice. Polycystic Ovary Syndrome. The New England journal of medicine 375, 54–64, https://doi.org/10.1056/NEJMcp1514916 (2016).
2. Dumesic, D. A. et al. Scientific Statement on the Diagnostic Criteria, Epidemiology, Pathophysiology, and Molecular Genetics of Polycystic Ovary Syndrome. Endocrine Reviews 36, 487–525, https://doi.org/10.1210/er.2015-1018 (2015).
3. Franks, S., Stark, J. & Hardy, K. Follicle dynamics and anovulation in polycystic ovary syndrome. Human reproduction update 14, 367–378, https://doi.org/10.1093/humupd/dmn013 (2008).
4. Kahsar-Miller, M. D., Nixon, C., Boots, L. R., Go, R. C. & Azziz, R. Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. Fertility and sterility 75, 53–58 (2001).
5. Vink, J. M., Sadrzadeh, S., Lambalk, C. B. & Boomsma, D. I. Heritability of polycystic ovary syndrome in a Dutch twin-family study. The journal of clinical endocrinology and metabolism 91, 2100–2104, https://doi.org/10.1210/jc.2005-1494 (2006).
6. Sam, S., Legro, R. S., Essah, P. A., Apridonidze, T. & Dunaif, A. Evidence for metabolic and reproductive phenotypes in mothers of women with polycystic ovary syndrome. Proceedings of the National Academy of Sciences of the United States of America 103, 7030–7035, https://doi.org/10.1073/pnas.0602025103 (2006).
7. Jansen, E. et al. Abnormal gene expression profiles in human ovaries from polycystic ovary syndrome patients. Molecular endocrinology (Baltimore, Md.) 18, 3050–3063, https://doi.org/10.1210/me.2004-0074 (2004).
8. Wood, J. R. et al. Valproate-induced alterations in human theca cell gene expression: clues to the association between valproate use and metabolic side effects. Physiological genomics 20, 233–243, https://doi.org/10.1152/physiogenomics.00193.2004 (2005).
9. Corton, M. et al. Differential gene expression profile in omental adipose tissue in women with polycystic ovary syndrome. The journal of clinical endocrinology and metabolism 92, 328–337, https://doi.org/10.1210/jc.2006-1665 (2007).
10. Azziz, R. PCOS in 2015: New insights into the genetics of polycystic ovary syndrome. Nature reviews. Endocrinology 12, 74–75, https://doi.org/10.1038/nrendo.2015.230 (2016).
11. Jones, M. R. & Goodarzi, M. O. Genetic determinants of polycystic ovary syndrome: progress and future directions. Fertility and sterility 106, 25–32, https://doi.org/10.1016/j.fertnstert.2016.04.040 (2016).
12. Liu, H., Zhao, H. & Chen, Z. J. Genome-Wide Association Studies for Polycystic Ovary Syndrome. Seminars in reproductive medicine 34, 224–229, https://doi.org/10.1055/s-0036-1585403 (2016).
13. Liu, H. Y., Huang, Y. L., Liu, J. Q. & Huang, Q. Transcription factormicroRNA synergistic regulatory network revealing the mechanism of polycystic ovary syndrome. Molecular medicine reports 13, 3920–3928, https://doi.org/10.3892/mmr.2016.5019 (2016).
14. Korti, M., Gov, E. & Arga, K. V. Molecular signatures of ovarian diseases: Insights from network medicine perspective. Systems biology in reproductive medicine 62, 266–282, https://doi.org/10.1080/19396368.2016.1197982 (2016).
15. Li, Y. H., Dong, M. Q. & Guo, Z. Systematic analysis and prediction of longevity genes in Caenorhabditis elegans. Mechanisms of ageing and development 131, 700–709, https://doi.org/10.1016/j.mad.2010.10.001 (2010).
16. Xu, J. & Li, Y. Discovering disease-genes by topological features in human protein–protein interaction network. Bioinformatics (Oxford, England) 22, 2800–2805, https://doi.org/10.1093/bioinformatics/btl467 (2006).
17. Lopez-Bigas, N. & Ouzounis, C. A. Genome-wide identification of genes likely to be involved in human genetic disease. Nucleic acids research 32, 3108–3114, https://doi.org/10.1093/nar/gkh665 (2004).
18. Li, Y. H., Zhang, G. G. & Wang, N. Systematic Characterization and Prediction of Human Hypertension Genes. Hypertension (Dallas, Tex.: 1979) 69, 349–355, https://doi.org/10.1161/hypertensionaha.116.08573 (2017).
19. Jeong, H., Mason, S. & Barabasi, A. Lethality and centrality in protein networks. Nature 411, 41–42 (2001).
20. Brown, K. R. & Jurisica, I. Online predicted human interaction database. Bioinformatics 21, 2076–2082, https://doi.org/10.1093/bioinformatics/btI273 (2005).
21. Wuchty, S. & Almaas, E. Peeling the yeast protein network. Proteomics 5, 444–449, https://doi.org/10.1002/pmic.200400962 (2005).
22. Yoon, J., Blumer, A. & Lee, K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics (Oxford, England) 22, 3106–3108, https://doi.org/10.1093/bioinformatics/btl533 (2006).
23. Furney, S. J., Higgins, D. G., Ouzounis, C. A. & Lopez-Bigas, N. Structural and functional properties of genes involved in human cancer. BMC genomics 7, 3, https://doi.org/10.1186/1471-2164-7-3 (2006).
24. Wu, X. Q. et al. The WNT/beta-catenin signaling pathway may be involved in granulosa cell apoptosis from patients with PCOS in North China. Journal of gynecology obstetrics and human reproduction 46, 93–99, https://doi.org/10.1016/j.jgor.2015.08.013 (2017).
25. Mbirek, H. et al. Identification of Common Genetic Variants Influencing Spontaneous Diysgotic Twinning and Female Fertility. American journal of human genetics 98, 898–908, https://doi.org/10.1016/j.ajhg.2016.03.008 (2016).
26. Denner, L., Bodenburg, Y. H., Jiang, J., Pages, G. & Urban, R. Insulin-like growth factor-I activates extracellularly regulated kinase to regulate the p450 side-chain cleavage insulin-like response element in granulosa cells. Endocrinology 151, 2819–2825, https://doi.org/10.1210/en.2009-1439 (2010).
27. Ozertan, K., Uncu, G. & Tuñeki, M. Insulin-like growth factor-1 and insulin-like growth factor-binding protein-1 in patients with polycystic ovary syndrome during clomiphene citrate therapy. International journal of gynecology and obstetrics: the official organ of the International Federation of Gynecology and Obstetrics 108, 71–72, https://doi.org/10.1016/j.ijgo.2009.08.016 (2010).
28. Zhong, G. & Chen, B. Serum and follicular fluid levels of IGF-II, IGF-binding protein-4 and pregnancy-associated plasma protein-A in controlled ovarian hyperstimulation cycle between polycystic ovarian syndrome (PCOS) and non-PCOS women. Gynecological endocrinology: the official journal of the International Society of Gynecological Endocrinology 27, 86–90, https://doi.org/10.101309/13590.2010.490608 (2011).
29. Welt, C. K. & Duran, J. M. Genetics of polycystic ovary syndrome. Seminars in reproductive medicine 32, 177–182, https://doi.org/10.1055/s-0033-1371089 (2014).
30. Urbanek, M. et al. Candidate gene region for polycystic ovary syndrome on chromosome 19p13.2. The Journal of clinical endocrinology and metabolism 90, 6623–6629, https://doi.org/10.1210/jc.2005-0622 (2005).
31. Myers, C. L., Barrett, D. R., Hibbs, M. A., Huttenhower, C. & Troyanskaya, O. G. Finding function: evaluation methods for functional genomic data. BMC genomics 7, 187, https://doi.org/10.1186/1471-2164-7-187 (2006).
32. Jee, W. C., Davis, S. E., Lapidus, L., Chain, E. D. & Hessler, R. D. The Journal of clinical endocrinology and metabolism 90, 6623–6629, https://doi.org/10.1210/jc.2005-0622 (2005).