
Application Notes

Phexpo: a package for bidirectional enrichment analysis of phenotypes and chemicals

Christopher Hawthorne,¹ David A. Simpson,¹ Barry Devereux² and Guillermo López-Campos¹

¹Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK and ²The Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast, UK

Corresponding Author: Guillermo López-Campos, PhD, Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK; g.lopezcampos@qub.ac.uk

Received 11 January 2020; Revised 13 March 2020; Editorial Decision 11 May 2020; Accepted 21 May 2020

ABSTRACT

Phenotypes are the result of the complex interplay between environmental and genetic factors. To better understand the interactions between chemical compounds and human phenotypes, and further exposome research we have developed “phexpo,” a tool to perform and explore bidirectional chemical and phenotype interactions using enrichment analyses. Phexpo utilizes gene annotations from 2 curated public repositories, the Comparative Toxicogenomics Database and the Human Phenotype Ontology. We have applied phexpo in 3 case studies linking: (1) individual chemicals (a drug, warfarin, and an industrial chemical, chloroform) with phenotypes, (2) individual phenotypes (left ventricular dysfunction) with chemicals, and (3) multiple phenotypes (covering polycystic ovary syndrome) with chemicals. The results of these analyses demonstrated successful identification of relevant chemicals or phenotypes supported by bibliographic references. The phexpo R package (<https://github.com/GHLCLab/phexpo>) provides a new bidirectional analyses approach covering relationships from chemicals to phenotypes and from phenotypes to chemicals.

Key words: exposome, phenotype, chemicals and drugs, computational biology/methods, biological ontologies

LAY SUMMARY

Chemicals are major contributors to the “exposome,” the whole set of exposures experienced by an individual that shapes their phenotype. Although the effects of exposure to some chemicals can be tested directly this is often not feasible. We conjectured that it might be possible to associate chemicals with phenotypes through the common genes to which they have been independently linked. In this manuscript we present “phexpo,” a novel tool that analyses the overlap between gene lists linked with chemicals or phenotypes. This approach enabled the detection of both known and novel associations between chemicals and phenotypes. Case studies using a chemical, a phenotype, or a combination of phenotypes as a query element demonstrate the application of phexpo. The relationships

identified between these entities were supported by evidence from the literature. Phexpo facilitates the establishment and discovery of novel relationships between chemicals and phenotypes and vice versa and therefore provides a valuable new tool for the study of the exposome.

BACKGROUND AND SIGNIFICANCE

Phenotypes, as described in the Human Phenotype Ontology (HPO),¹ are the result of the complex interplay between environmental and genetic factors. Recognition of the importance of environmental factors, coupled with an increasing ability to determine individual exposures, led to the concept of the “exposome” as the

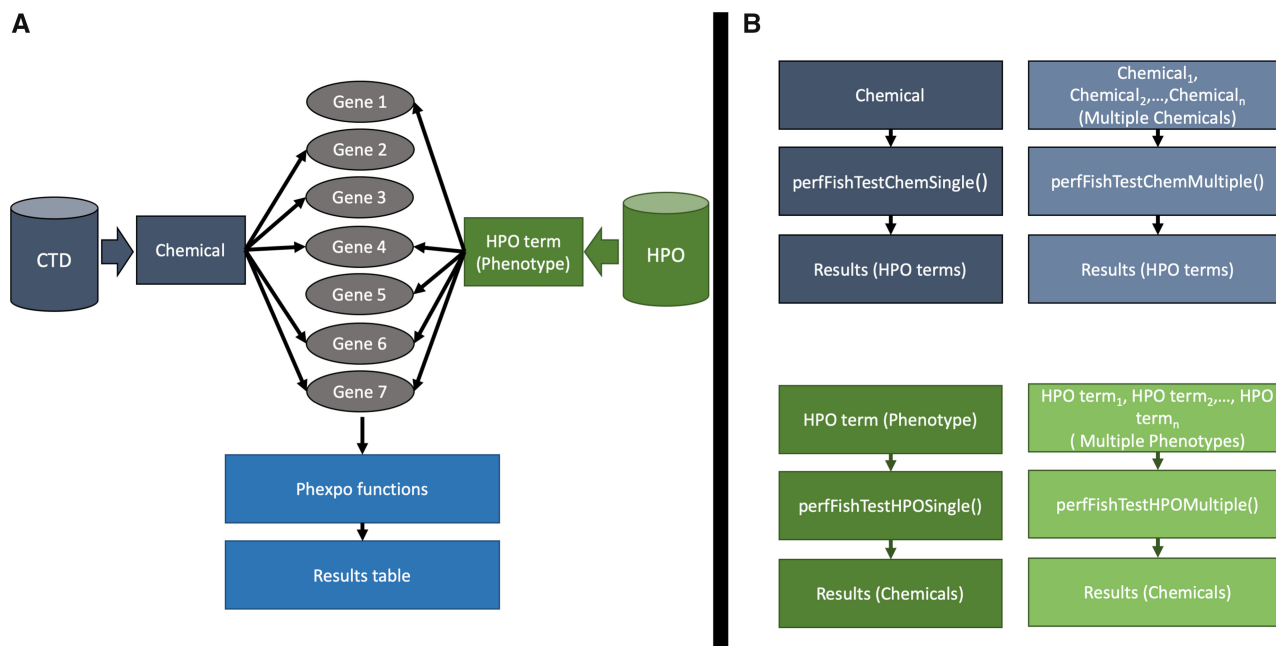


Figure 1. Diagrammatic representation of phexpo's processes. (A) Chemicals and phenotypes can be connected via genes. Phexpo's analytical functions return a table of associated results. (B) Further breakdown of phexpo functions. If a user inserts a chemical into the perfFishTestChem functions enriched phenotypes are returned. Conversely, if a user inserts a phenotype into the perfFishTestHPO functions enriched chemicals are returned.

whole set of exposures of an individual since conception.² The exposome soon gained traction in research as a complement to the genome and has become an important element for the development of new precision medicine applications.³

During the last century chemistry revolutionized many industries and human activities and as a result, the environment now contains an unprecedented amount and variety of chemicals, such as plasticizers used in bottles, flame retardants in clothes, new drugs, and pesticides used to improve crop yields.⁴ This increase in anthropogenic chemicals has fostered advances in toxicology and monitoring, and the plethora of subsequent studies and data are making the “chemical component” one of the best understood and studied elements of the exposome. However, despite some advances, our understanding of the biological effects of most chemical compounds is incomplete and for many there remains significant controversy regarding safety levels and effects upon human health.⁵

Biomedical informatics and translational bioinformatics provide analytical tools and develop data repositories to support exposome and toxicological research. Some of these repositories, such as Toxin and Toxin Target Database,⁶ Exposome Explorer,⁷ or the Comparative Toxicogenomics Database (CTD),⁸ integrate different data sources and combine chemical and biological information such as biomarkers or target genes. This integration has enabled the development of analytical approaches using these contents and the relationships between chemicals and genes to uncover potential links between chemicals and biological pathways or diseases. Although these approaches provide valuable insights, Gene Ontology terms and pathways are mostly focused on the biological outcomes at a cellular or molecular level whereas analyses of diseases involve sets of phenotypes. There is therefore an unmet need to implement a method to relate chemical compounds to the different phenotypes described in the HPO. The HPO has formalized the phenotype space by providing descriptions of clinical abnormalities and annotations to both rare and common diseases and is increasingly being used by

different actors for data exchange and identification of disease etiology. To facilitate a bidirectional analysis of the relationships between chemicals and phenotypes using gene annotations, we have developed phexpo (phenotype–exposome), a methodology to perform bidirectional enrichment analysis of chemicals and phenotypes. This methodology has been bundled inside an R package. Phexpo incorporates chemical and gene data from CTD and phenotype and gene data from HPO. We refer to phenotypes as HPO terms. Using a chemical- or phenotype-derived gene list built from genes in both data sources, phexpo will provide enriched chemicals or phenotypes, respectively. This demonstrates a novel methodology that combines gene information from CTD and HPO to generate potential associations between chemical exposures and phenotypes.

METHODS

Chemical–gene relationships, chemical vocabulary, and phenotype–gene datasets were downloaded from CTD (February 5, 2019 update) and HPO (ontology version: February 12, 2019), respectively. As our focus is on human phenotypes, CTD datasets were preprocessed to concentrate exclusively on human gene identifiers and generate the relevant gene lists linking chemicals and human genes. A more detailed preprocessing explanation is included in the [Supplementary Material](#).

To identify relationships between chemicals and phenotypes, phexpo uses the gene lists derived from the annotations for chemicals and phenotypes and then compares them using a Fisher's exact test against a background universe of genes generated from the aforementioned annotations. To correct for multiple testing phexpo includes and reports Bonferroni corrected *P*-values and a false discovery rate using Benjamini and Hochberg corrected *P*-values.

Phexpo is built around 4 analytical functions (Figure 1) and 1 visualization function:

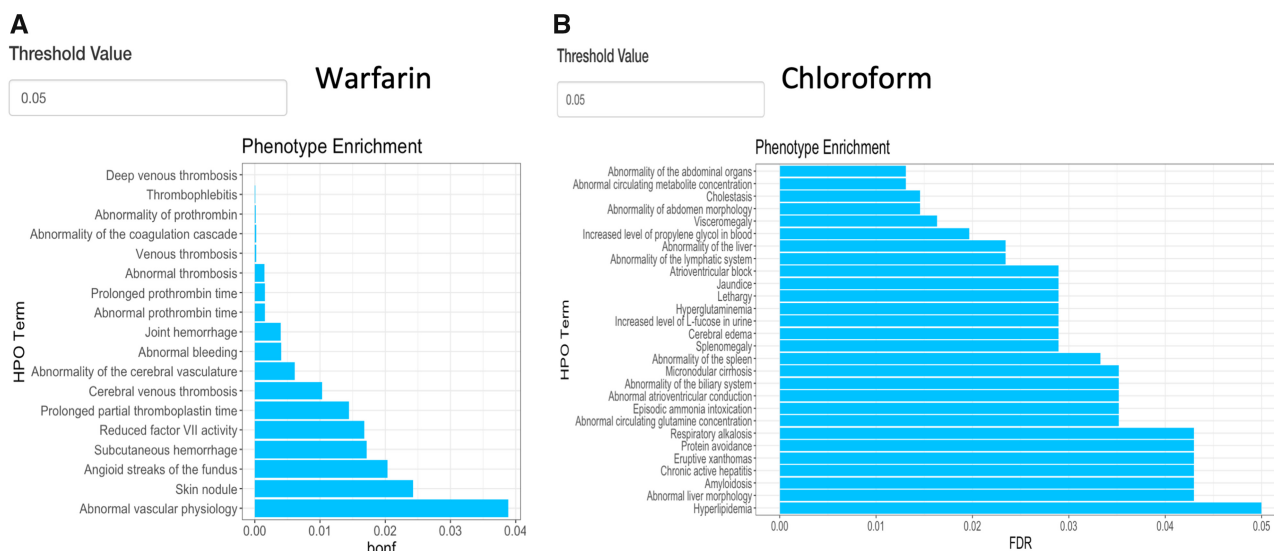


Figure 2. Case study I results using the shiny interface. The bar charts interface enables filtering using different criteria (A) shows HPO terms identified for warfarin filtered by Bonferroni correction. (B) HPO terms identified for chloroform filtered by FDR. Full results tables available in the [Supplementary Material](#). *Abbreviations:* FDR: false discovery rate; HPO: Human Phenotype Ontology.

1. `perffishTestChemSingle()`: This function uses a chemical name provided by the user as input and generates a gene list using the chemical–gene dataset. It uses only the genes that are in the intersection of the chemical–gene and phenotype–gene datasets. It calculates the various gene counts against the phenotype–gene dataset and uses them for the R built-in Fisher’s exact test. The function returns a table with all the associated phenotypes.
2. `perffishTestHPOSingle()`: This function uses a HPO term as input and carries out a calculation analogous to `perffishTestChemSingle()`, but utilizes the phenotype–gene dataset for gene list creation. It calculates gene counts against the chemical–gene dataset to run the Fisher’s exact test and return a table with all the associated chemicals.
3. `perffishTestHPOMultiple()`: This function uses a list of different phenotypes. It aggregates all the annotated genes for the given phenotypes into a single gene list for the comparison and then carries out the same role as `perffishTestHPOSingle()` to return a table with all the associated chemicals.
4. `perffishTestChemMultiple()`: This function uses a list of different chemicals. It aggregates all the annotated genes for the given chemicals into a single gene list for the comparison and then carries out the same role as `perffishTestChemSingle()` to return a table with all the associated phenotypes.
5. `visEnrich()`: This function provides a Shiny⁹ interface for the visualization of the results generated using any of the other 4 analytical functions. This function generates a graphical user interface that presents the results in a tabular format and a graphical display enabling the user to manipulate the results using different filtering criteria.

The results of the 4 analytical functions are presented in tabular format and include the raw and corrected *P*-values, the different gene set sizes, and their overlaps.

RESULTS

To facilitate the bidirectional integration of chemicals and phenotypes we developed a new approach built into an R package, “phexpo,” that

exploits gene annotations extracted from 2 curated high-quality resources, CTD (for chemical–gene annotations) and HPO (for phenotype–gene annotations). To demonstrate the capabilities of phexpo we present 3 different case studies using its different functions. In order to assess and evaluate the results we manually validated some of the results and provide bibliographic evidence supporting the associations (additional details are in the [Supplementary Material](#)).

Case study I—single chemical to phenotype enrichment

To validate the analysis of single chemicals in phexpo we chose 2 diverse, but well-studied compounds with predictable results, a drug (warfarin) and an industrial chemical (chloroform) ([Figure 2](#)). For this analysis, we used the function `perffishTestChemSingle()`. As expected, the enriched phenotypes for warfarin, including “deep vein thrombosis” and “abnormality of prothrombin,” match its anticoagulant function. For chloroform, we identified liver phenotypes consistent with its known hepatotoxicity.¹⁰ Additional analysis can be found in [Supplementary Material](#).

Case study II—single phenotype to chemical enrichment

For the single phenotype case study, we used the HPO term “left ventricular dysfunction” which is described as “inability of the left ventricle to perform its normal physiologic function. Failure is either due to an inability to contract the left ventricle or the inability to relax completely and fill with blood during diastole.” For this analysis, we used the function `perffishTestHPOSingle()` and the top 10 results are presented in [Table 1](#).

From these results we highlight the following potential relationships, halofuginone has been found to elicit a shielding effect against stress on the heart.¹¹ 1-Trifluoromethoxyphenyl-3-(1-propionylpiperidine-4-yl) urea is a soluble epoxide hydrolase inhibitor and soluble epoxide hydrolase inhibitors have been suggested as a potential strategy against heart diseases.¹² In animal models streptozocin is used to cause diabetes mellitus, which has a link to diastolic heart dysfunction.¹³ The antibiotic bleomycin has known cardiotoxic effects when used in chemotherapy.¹⁴ Fenofibrate given short term

Table 1. Top 10 chemicals identified for the HPO term “left ventricular dysfunction”

Chemical name	P-value	Bonf	FDR
Halofuginone	1.11E-11	1.31E-08	1.31E-08
Nitrofen	6.15E-11	7.22E-08	3.61E-08
1-Trifluoromethoxyphenyl-3-(1-propionylpiperidine-4-yl) urea	1.04E-08	1.22E-05	3.54E-06
Streptozocin	1.21E-08	1.42E-05	3.54E-06
Bleomycin	2.03E-08	2.39E-05	4.07E-06
Fenofibrate	2.08E-08	2.44E-05	4.07E-06
Phenylephrine	3.33E-08	3.91E-05	5.59E-06
Palm Oil	6.59E-08	7.74E-05	9.48E-06
Doxorubicin	9.25E-08	0.000109	9.48E-06
Dietary fats	9.65E-08	0.000113	9.48E-06

Arranged by ascending P-value.

Abbreviations: HPO: Human Phenotype Ontology; FDR: false discovery rate.

Table 2. Top 10 chemical results identified for PCOS phenotypes

Chemical name	P-value	Bonf	FDR
Tetrachlorodibenzodioxin	2.38E-39	1.22E-35	1.22E-35
Bisphenol A	2.70E-32	1.39E-28	6.94E-29
Ammonium chloride	1.21E-25	6.23E-22	2.08E-22
Valproic acid	1.50E-19	7.71E-16	1.79E-16
Ethylnitrosourea	1.74E-19	8.93E-16	1.79E-16
Colforsin	4.17E-19	2.15E-15	3.58E-16
Vehicle emissions	6.76E-19	3.48E-15	4.96E-16
Diethylhexyl phthalate	2.55E-18	1.31E-14	1.64E-15
Ethinyl estradiol	3.52E-18	1.81E-14	2.01E-15
Dexamethasone	3.20E-17	1.65E-13	1.65E-14

Arranged by ascending P-value.

Abbreviations: FDR: false discovery rate; PCOS: polycystic ovary syndrome.

has been shown to reduce some of the effects of chronic left ventricular volume overload in rat models.¹⁵ Phenylephrine causes hypertrophy when administered to neonatal rat cardiomyocytes.¹⁶ Tocotrienol rich fractions extracted from palm oil had beneficial heart functioning.¹⁷ Doxorubicin is both a drug used for cancer treatment as well as being a cardiotoxic agent involved in causing heart failure.¹⁸

Case study III—multiple phenotypes to chemical enrichment

An important feature of our methodology is that it allows for the combination of multiple phenotypes in a single enrichment analysis. This is important for diseases, complex conditions, or syndromes that comprise a variety of phenotypes, which can be stacked together in our analysis. In this case study, we use the example of polycystic ovary syndrome (PCOS). Although the wide range of different phenotypes displayed by women with this endocrine disorder have hindered elucidation of the causes, exosomic and inherited genetic variables are likely to play a part.¹⁹

We compiled a list of phenotypes including “oligomenorrhea,” “enlarged polycystic ovaries,” “amenorrhea,” “hirsutism,” “increased body weight,” and “acne” that characterize PCOS^{20,21} and performed an enrichment analysis to test whether it returns chemicals with a known relationship to PCOS shown in Table 2.

Of the multiple highly enriched chemicals returned, many have documented links with PCOS. Tetrachlorodibenzodioxin is an endocrine-disrupting chemical and although its influence on PCOS has not been specifically assessed it has been highlighted as a suspect for consideration.²² Individuals with PCOS were found to have increased levels of bisphenol A.²³ Valproic acid treatment has a known association with heightened PCOS occurrence in epileptic patients.²⁴ Colforsin (forskolin) can function in a similar way to luteinizing hormone on PCOS theca cells.²⁵ Diethylhexyl phthalate has been shown to have ovarian effects in rats.²⁶ Ethinyl estradiol is used to combat the acne and hirsutism phenotypes of PCOS.²⁷ Dexamethasone has been used to increase testosterone production in theca cells to mimic PCOS patients with hyperandrogenism.²⁸

DISCUSSION

In this work, we have presented a new approach to establish bidirectional relationships between phenotypes and chemical exposures. This methodology has been successfully implemented in phexpo, a multiplatform R package that is freely available (<https://github.com/GHLCLab/phexpo>).

In contrast to other existing applications our approach allows searches with multiple chemicals or phenotypes simultaneously. This enables users to search (or find) different individual phenotypes rather than diseases that might be too broad or may share phenotypes (or symptoms) that could lead to overlaps.

We have successfully tested phexpo and its different functions using a variety of chemicals and phenotypes and have been able to validate the results using bibliographic references. The methodology can therefore be used to discover novel potential relationships that open new avenues of research and direct additional experimental validations and exploration.

Although we demonstrate the ability of the approach to generate interesting and validated results, we acknowledge that there are limitations. The results are confined to the annotations present and although we selected high-quality and well-known resources we are currently only using 1 chemical database (CTD) and 1 phenotype ontology (HPO). The associations established between chemicals and phenotypes lack “directionality,” in that a chemical may induce or protect from a certain phenotype, and indeed both were found in the case study lists. Finally, although aggregating phenotypes or chemicals is a powerful tool, the use of an additive approach requires the union of all the annotated genes and does not take account of whether in 2 different phenotypes or chemicals those genes might be affected in different directions (eg, induced by 1 chemical and repressed by another). Other potential limitations are that dosage and timing of exposures that might be relevant for the development of some of the phenotypes are not considered.

In conclusion, we have introduced a novel methodology bundled inside an R package called phexpo that links chemical compounds and phenotype terms through enrichment analyses based on their gene annotations. We have described 3 case studies validated through the literature which present phexpo’s functionalities and its capabilities to identify phenotypes related to a chemical and vice versa. Phexpo’s bidirectional approach to study the potential relationships between chemical compounds and human phenotypes provides insights for human health and exposome research. This tool will be a valuable asset to further exposome research by revealing potential novel phenotype–chemical associations.

FUNDING

CH has been supported by a Northern Ireland Department for the Economy (DfE) postgraduate studentship award.

AUTHOR CONTRIBUTIONS

CH constructed the package, contributed to the idea, contributed to the design of the experiments, run the analyses, validated the results, and wrote the manuscript. GL-C created the idea, designed the experiments, validated the results, and wrote the manuscript. DAS contributed to the idea and critical manuscript revision. BD contributed to the idea and critical manuscript revision.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGEMENTS

We thank Dr Jaine Blayney for reading a draft version of the manuscript and the useful comments and feedback returned.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Köhler S, Carmody L, Vasilevsky N, *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* 2019; 47 (D1): D1018–27.
- Wild CP. Complementing the genome with an ‘exposome’: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005; 14 (8): 1847–50.
- Niedzwiecki MM, Walker DI, Vermeulen R, *et al.* The exposome: Molecules to populations. *Annu Rev Pharmacol Toxicol* 2019; 59 (1): 107–27.
- Dong T, Zhang Y, Jia S, *et al.* Human indoor exposome of chemicals in dust and risk prioritization using EPA’s ToxCast database. *Environ Sci Technol* 2019; 53 (12): 7045–54.
- Prins GS, Patisaul HB, Belcher SM, *et al.* CLARITY-BPA academic laboratory studies identify consistent low-dose bisphenol A effects on multiple organ systems. *Basic Clin Pharmacol Toxicol* 2019; 125 (S3): 14–31.
- Wishart D, Arndt D, Pon A, *et al.* T3DB: The toxic exposome database. *Nucleic Acids Res* 2015; 43 (D1): D928–34.
- Neveu V, Moussy A, Rouaix H, *et al.* Exposome-explorer: A manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res* 2017; 45 (D1): D979–84.
- Davis AP, Grondin CJ, Johnson RJ, *et al.* The comparative toxicogenomics database: Update 2019. *Nucleic Acids Res* 2019; 47 (D1): D948–54.
- Chang W, Cheng J, Allaire JJ, *et al.* shiny: Web Application Framework for R. 2018. <https://cran.r-project.org/web/packages/shiny/index.html> Accessed June 10, 2020.
- Minor EA, Newman MS, Kupec JT. Chronic, recreational chloroform-induced liver injury. *Case Rep Hepatol* 2018; 2018: 1–3.
- Qin P, Arabacilar P, Bernard RE, *et al.* Activation of the amino acid response pathway blunts the effects of cardiac stress. *J Am Heart Assoc* 2017; 6: e004453.
- Guo Y, Luo F, Zhang X, *et al.* TPPU enhanced exercise-induced epoxyeicosatrienoic acid concentrations to exert cardioprotection in mice after myocardial infarction. *J Cell Mol Med* 2018; 22 (3): 1489–500.
- Sorop O, Heinonen I, van Kranenburg M, *et al.* Multiple common comorbidities produce left ventricular diastolic dysfunction associated with coronary microvascular dysfunction, oxidative stress, and myocardial stiffening. *Cardiovasc Res* 2018; 114 (7): 954–64.
- Madeddu C, Deidda M, Piras A, *et al.* Pathophysiology of cardiotoxicity induced by nonanthracycline chemotherapy. *J Cardiovasc Med (Hagerstown)* 2016; 17: e12–8.
- Dhahri W, Couet J, Roussel É, *et al.* Fenofibrate reduces cardiac remodeling and improves cardiac function in a rat model of severe left ventricle volume overload. *Life Sci* 2013; 92 (1): 26–34.
- Zhang B, Zhang P, Tan Y, *et al.* C1q-TNF-related protein-3 attenuates pressure overload-induced cardiac hypertrophy by suppressing the p38/CREB pathway and p38-induced ER stress. *Cell Death Dis* 2019; 10 (7): 520.
- Wong W-Y, Poudyal H, Ward LC, *et al.* Tocotrienols reverse cardiovascular, metabolic and liver changes in high carbohydrate, high fat diet-fed rats. *Nutrients* 2012; 4 (10): 1527–41.
- El-Naggar AE, El-Gowilly SM, Sharabi FM. Possible ameliorative effect of ivabradine on the autonomic and left ventricular dysfunction induced by doxorubicin in male rats. *J Cardiovasc Pharmacol* 2018; 72 (1): 22–31.
- Zhu X, Hong X, Chen L, *et al.* Association of methylenetetrahydrofolate reductase C677T and A1298C polymorphisms with genetic susceptibility to polycystic ovary syndrome: A PRISMA-compliant meta-analysis. *Gene* 2019; 719: 144079.
- The Human Phenotype Ontology. Polycystic Ovary Syndrome 1 OMIM: 184700. 2019. <https://hpo.jax.org/app/browse/disease/OMIM:184700> Accessed August 19, 2019.
- NHS. Polycystic Ovary Syndrome. 2019. <https://www.nhs.uk/conditions/polycystic-ovary-syndrome-pcos/> Accessed September 3, 2019.
- Merkin SS, Phy JL, Sites CK, *et al.* Environmental determinants of polycystic ovary syndrome. *Fertil Steril* 2016; 106 (1): 16–24.
- Konieczna A, Rachoń D, Owczarek K, *et al.* Serum bisphenol A concentrations correlate with serum testosterone levels in women with polycystic ovary syndrome. *Reprod Toxicol* 2018; 82: 32–7.
- Sidhu HS, Srinivasa R, Sadhotra A. Evaluate the effects of antiepileptic drugs on reproductive endocrine system in newly diagnosed female epileptic patients receiving either valproate or lamotrigine monotherapy: A prospective study. *Epilepsy Res* 2018; 139: 20–7.
- Wood JR, Nelson VL, Ho C, *et al.* The molecular phenotype of polycystic ovary syndrome (PCOS) theca cells and new candidate PCOS genes defined by microarray analysis. *J Biol Chem* 2003; 278 (29): 26380–90.
- Davis BJ, Maronpot RR, Heindel JJ. Di-(2-ethylhexyl) phthalate suppresses estradiol and ovulation in cycling rats. *Toxicol Appl Pharmacol* 1994; 128 (2): 216–23.
- Ruan X, Kubba A, Aguilar A, *et al.* Use of cyproterone acetate/ethinylestradiol in polycystic ovary syndrome: Rationale and practical aspects. *Eur J Contracept Reprod Health Care* 2017; 22 (3): 183–90.
- Ong M, Cheng J, Jin X, *et al.* Paeoniflorin extract reverses dexamethasone-induced testosterone over-secretion through downregulation of cytochrome P450 17A1 expression in primary murine theca cells. *J Ethnopharmacol* 2019; 229: 97–103.