**Review Article**

**Understanding the Ingenuity Pathway Analysis Software for Omics Research**

**Samantha A. Cintron, BSN, RN, PCCN1#, Qiuhua Shen, PhD, APRN1, Diane Mahoney, PhD, DNP, FNP-BC, WHNP-BC1, Mihaela E. Sardiu, PhD2, John B. Hiebert, MD1, Janet Pierce, PhD, APRN, CCRN, FAAN1**

1School of Nursing, University of Kansas Medical Center, Kansas, USA

2Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas, USA

**#Corresponding author:** Samantha A. Cintron, BSN, RN, PCCN, PhD Student, School of Nursing, University of Kansas Medical Center, Mail Stop 4043, 3901 Rainbow Blvd, Kansas City, Kansas 66160, USA

**How to cite this article:** Cintron SA, Shen Q, Mahoney D, et al. (2021) Understanding the Ingenuity Pathway Analysis Software for Omics Research. Int J Nurs & Healt Car Scie 01(15): 2021-72.

**Submission Date:** 18 July, 2021; **Accepted Date:** 29 October, 2021; **Published Online:** 04 November, 2021

**Abstract**

**Aim**: The overall aim of this article was to provide a brief review of the Ingenuity Pathway Analysis software used to examine omics data. In this article, we reviewed the five major core analyses used to understand omics data within the context of various biological systems using omics data from a Traumatic Brain Injury (TBI) experiment.

**Background**: Qiagen Ingenuity Pathway Analysis (IPA) is a powerful web-based software program used by researchers and clinicians to visualize, analyze, and understand complex omics data. This program is built on biomedical databases derived over 20 years to assist with the integration and aggregation of various signaling and metabolic pathways.

**Design**: We demonstrated the application of IPA using cerebral omics data generated from experiments where we investigated the effects of ubiquinol on gene expression in rats following TBI.

**Method**: We applied IPA to understand the significant pathways, upstream regulators, biological functions, and interaction networks specifically related to TBI when administering ubiquinol.

**Conclusion**: Large amount of omics data were generated from a small number of brain tissues (n = 4). Despite a small sample size, we found significant functional and mechanistic changes. Using IPA allowed us to identify the key networks, causal relationships, and novel regulatory pathways to assist with discovery of potential biomarkers.

**Keywords:** Bioinformatics; Biomarkers; Data science; Gene expression; Ingenuity pathway analysis

**What is known?**

* IPA is a web-based bioinformatics application used to understand omics type data.
* This software program has assisted researchers and clinicians to upload omics data results from genomic, metabolomic, proteomic, and other omics experiments.

**What is new?**

* IPA has a search capability for information on genes, proteins, chemicals, and drugs and allows interactive building of networks to represent biological systems.
* IPA can explore omics data for individual genes/proteins and detect change of expression patterns between compared samples.
* There are robust visualizations of networks and pathways that can be used to better understand disease processes to improve clinical outcomes.
* These powerful analyses and search tools can reveal the implication of data and help recognize new targets or candidate biomarkers for specific diseases.

**Introduction**

With an increasing focus on precision health and omics based research, it is essential to have advanced software to analyze omics data to assist with understanding biological mechanisms involved in various diseases [1]. After the completion of the Human Genome Project, there were massive data repositories that needed an advanced analytical software program to assist with analysis, integration, and understanding of the data from gene expression, microRNA (miRNA), and Single Nucleotide Polymorphisms (SNPs) [2,3]. One program that has been used by most omics researchers is called Ingenuity Pathway Analysis (IPA) [4]. This leading pathway analysis software application program performs advanced pathway analysis and interprets results while applying known biological relationships. IPA is a web-based bioinformatics application where these biological (causal and directional) relationships are compiled from millions of manually reviewed findings by PhD-prepared curators to form the Ingenuity Knowledge Base [4,5]. The Ingenuity Knowledge Base is updated on a weekly and quarterly basis to ensure that the most recent findings are included. Direct links to the original source articles are integrated within the program for easy access [6,7].

IPA can be used for analysis with any size of experimental data and allows the researcher or clinician to target specific information concerning drugs, proteins, genes, and chemicals. Thus, data generated from experiments using proteomics, genomics, transcriptomics, metabolomics, and lipidomics can be used to determine downstream effects, upstream regulators, and identify new targets for biomarkers. Powerful algorithms are used to identify the most significant pathways and whether there is activation or inhibition of novel regulatory networks with causal relationships [8].

There are courses and many web-based training programs available related to IPA [4]. For this article, we will review the 5 major IPA core analysis categories that are considered data-derived networks and illustrate the use of the IPA core analyses in a transcriptomics study on Traumatic Brain Injury (TBI). Description of the methods and primary analysis can be found in the original article [9]. Data were analyzed and figures were developed using IPA (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). These five core analyses are: (1) Canonical pathways, (2) Upstream regulators, (3) Diseases and functions, (4) Regulatory effects, and (5) Networks (Table 1) [5]. IPA core analyses assist the researcher and clinician to identify functions, pathways, relationships, and mechanisms relevant to disease. The purpose of this article is to provide clinicians and researchers with an introductory overview of the functions in IPA to increase the understanding of how this complex analysis software can be used to connect phenotypes with the biological basis of disease.

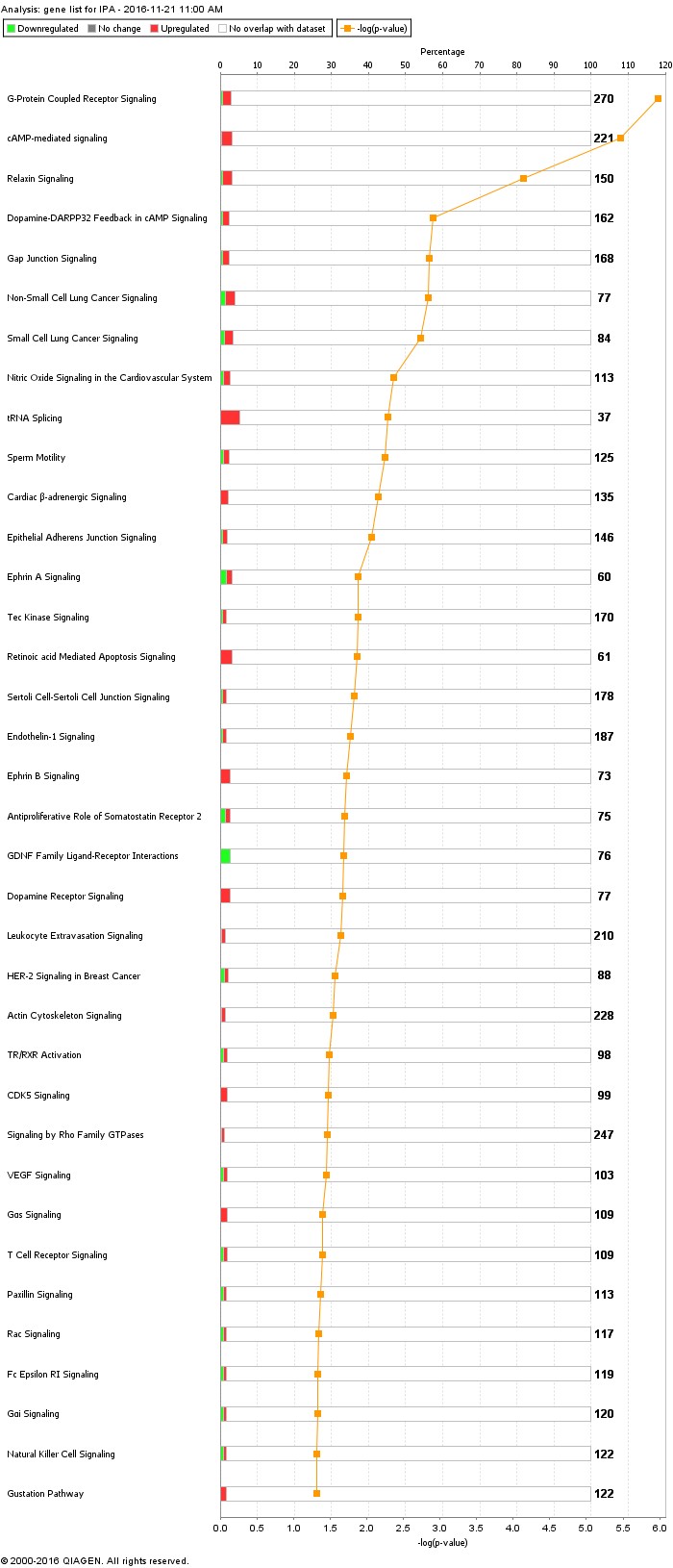
|  |  |
| --- | --- |
| **Categories** | **Function** |
| 1. Canonical Pathways | Predicts pathways that are changing based on dataset Predicts directional effects on the pathway molecules not in dataset (MAP overlay tool) |
| 2. Upstream Regulators | Predicts activated/inhibited regulators responsible for observed data Predicts master regulators |
| 3. Diseases and Functions | Predicts the directional biological effects (cellular processes, biological function) of gene/protein set Examples include: • Increase in cell cycle • Decrease in apoptosis |
| 4. Regulator Effects | Identifies specific hypothesis: upstream regulator pathways leading to a downstream phenotype. |
| 5. Networks | Identifies gene networks within dataset |
| *Note:* MAP = Molecule Activity Predictor | |

**Table 1:** IPA core analysis categories with function.

**Canonical Pathways**

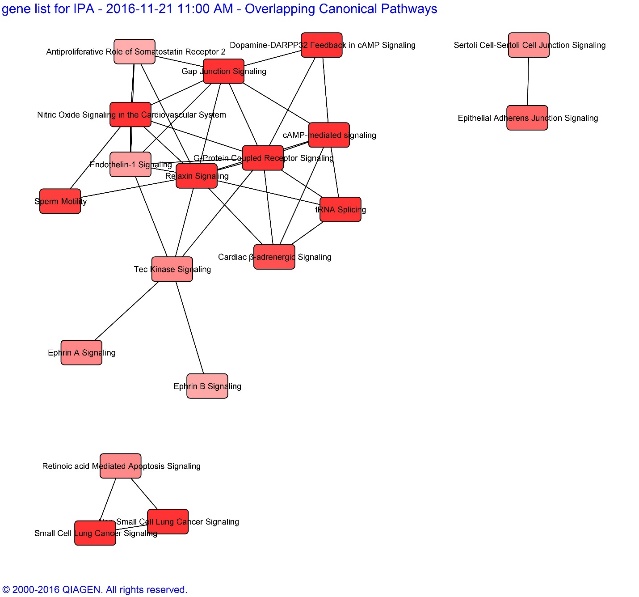
Canonical pathways are signaling and metabolic pathways that are based on the literature and well established within the scientific community [5,10]. Because canonical pathways are established, they do not change with input data. After experimental input data are analyzed, the differentially expressed molecules and genes found in the dataset are compared to the canonical pathways [11,12]. Statistical analysis is used to identify the most enriched pathways and to predict if the canonical pathway is activated or inhibited [13,14]. Two statistical tests are used in IPA: Fisher’s exact test and the z-score. Fisher’s exact test determines the p-value of overlap that is used in all analyses. If the p-value is significant (i.e. p-value<0.05), there is overlap of the molecules in the dataset with a particular disease, function, or pathway that is not due to chance alone. The z-score, which represents the number of standard deviations from the mean of a normal distribution, is used in canonical pathways, upstream analysis, and diseases and functions analysis to determine if a molecule is activated or inhibited. An activated canonical pathway is determined by a positive z-score (≥ ± 2) and an inhibited pathway is determined by a negative z-score (≤ ± 2). If no prediction can be made about activation or inhibition from the current information in the Ingenuity Knowledge Base, the canonical pathway is labeled as having no activity pattern available [5].

(Figure 1) illustrates the top ten canonical pathways of differentially expressed genes in the brain of rats with Traumatic Brain Injury (TBI) that were treated with ubiquinol compared to an untreated group. The number of molecules from the dataset that are part of the canonical pathway are identified on the right side of the figure. For example, the relaxin signaling canonical pathway is composed of 150 molecules with five molecules identified in the dataset. One gene is downregulated (green), and four genes are upregulated (red). Relaxin mediates many biological activities including decreasing fibrosis, vasodilation, blood vessel formation, inhibiting histamine release, reducing apoptosis, and enhancing the release of other neuropeptides [15].IPA also provides an overlap ratio, which refers to the number of molecules that have been identified as part of the canonical pathway in the dataset divided by the total number of molecules in the canonical pathway. If a significant number of the molecules from dataset are found in the pathway, it could suggest that the canonical pathway is important. The overlap ratio of the dataset with the relaxin signaling canonical pathway was 3.3% [9].



**Figure 1:** Canonical Pathways of Differentially Expressed Genes in the Brain of Rats with TBI that are Ubiquinol-Treated Compared to Untreated.

IPA also provides a diagram of the canonical pathway that includes the location of the molecules in the cell, how the molecules interact with each other, and if the molecules are activated or inhibited [12]. Using the Molecule Activity Predictor (MAP) overlay tool, molecules can be selected by the user to be activated or inhibited. Directional effects on the pathway are predicted based on known relationships from the Ingenuity Knowledge Base. Another function of IPA is comparing multiple canonical pathways to identify common genes using the overlap function (Figure 2). An overlay can be selected to show the number of common genes between canonical pathways [5].



**Figure 2:** Overlapping Canonical Pathways of Differentially Expressed Genes in the Brain of Rats with TBI that are Ubiquinol-Treated Compared to Untreated. Pathway scores are displayed using a red color gradient, where darker red corresponds to higher scores (increased statistical significance).

**Upstream Regulators**

In IPA, any molecule that has a downstream effect on expression, transcription, protein-DNA binding, or phosphorylation is considered an upstream regulator [8]. Examples of upstream regulators include genes, transcription factors, miRNAs, and drugs [14]. Each upstream regulator that could be responsible for the expression changes found in the dataset is identified using the Ingenuity Knowledge Base [16]. The predicted activation status is calculated based on enrichment of downstream molecules to determine if the upstream regulator was activated or inhibited in the dataset [13]. The upstream regulator analysis can help predict molecules that are regulating genes in an experiment providing more information on the biological activities [13,14]. A network diagram of multiple upstream regulators can be used to understand how the upstream regulators interact and help predict master regulators [16].

In a pilot study of differentially expressed genes after ubiquinol (reduced form of CoQ10) was used as a treatment for TBI in rats, the top upstream regulators were identified as Huntingtin Gene (HTT), levodopa (L-dopa), Kruppel-like factor 16 (KLF16), Beta-estradiol, and microRNA 543-3p (miR-543-3p). L-dopa was the only upstream regulator predicted to be activated based on the enrichment states of the downstream molecules [9].

**Diseases and Functions**

IPA evaluates the differences between intervention and control groups to identify the diseases, biofunctions, and toxicity functions that are associated as downstream effects of the differentiated molecules in the dataset. The number of molecules present in the dataset associated with the disease or function is identified and compared to the dataset using the p-value of overlap [17]. A heat map is a visual representation of which diseases and biological functions are significantly enriched based on the molecules found in the dataset [12]. The information from Ingenuity Knowledge Base is used to predict the directional biological effects of the full set of molecules. This determines if there is an increase or decrease in a specific disease or function [18]. Toxicity functions are similar to the information found in top diseases and biofunctions, but focus on cardiotoxicity, hepatotoxicity, and nephrotoxicity [19]. Experimental data are linked to clinical pathology endpoints, which enables an understanding of pharmacological response and support mechanism-of-action hypothesis generation. Associated biomarkers for downstream effects of diseases, biofunctions, and toxicity functions could be selected for further study from this information [20].

A summary of the top diseases and biofunctions from the TBI study are listed in (Table 2). The top diseases identified in the study of ubiquinol treated rats with TBI included hereditary disorder (26 molecules), neurological disease (31 molecules), organismal injury and abnormalities (60 molecules), psychological disorders (24 molecules), and skeletal and muscular disorders (26 molecules). The molecular and cellular functions that were identified included nucleic acid metabolism (12 molecules), small molecule biochemistry (16 molecules), cellular function and maintenance (19 molecules), molecular transport (21 molecules), and cell signaling (12 molecules). The top physiological system development and function includes behavior (18 molecules), nervous system development and function (25 molecules), tissue development (26 molecules), digestive system development and function (14 molecules), and organismal development (19 molecules). Clinical chemistry showed increased levels of creatinine, potassium, and hematocrit, although each of these functions only had one molecule related to the increase [9]. Results for significant values should be critically analyzed to determine the most impactful downstream effects.

|  |  |  |
| --- | --- | --- |
| **Diseases and Disorders** | **p-value** | **#Molecules** |
| HTT | 8.64E-03 - 3.92E-09 | 26 |
| L-dopa | 9.28E-03 - 3.92E-09 | 31 |
| Klf16 | 9.68E-03 - 3.92E-09 | 60 |
| Beta-estradiol | 8.85E-03 - 3.92E-09 | 24 |
| miR-543-3p | 8.64E-03 - 3.92E-09 | 26 |
| **Molecular and Cellular Functions** | | |
| Nucleic Acid Metabolism | 6.88E-03 - 2.72E-07 | 12 |
| Small Molecule Biochemistry | 9.56E-03 - 2.72E-07 | 16 |
| Cellular Function and Maintenance | 9.19E-03 - 4.56E-07 | 19 |
| Molecular Transport | 9.56E-03 - 1.31E-06 | 21 |
| Cell Signaling | 9.19E-03 - 7.79E-06 | 12 |
| **Physiologic System Development and Function** | | |
| Behavior | 8.64E-03 - 4.44E-08 | 18 |
| Nervous System Development and Function | 8.85E-03 - 4.56E-07 | 25 |
| Tissue Development | 9.56E-03 - 4.56E-07 | 26 |
| Digestive System Development and Function | 8.64E-03 - 1.31E-05 | 14 |
| Organismal Development | 9.56E-03 - 4.75E-05 | 19 |
| *Note.* HTT = huntingtin gene; L-dopa = levodopa; KLF16 = Kruppel-like factor 16; miR-543-3p = microRNA 543-3p | | |

**Table 2:** The Top Diseases and Functions of Differentially Expressed Genes in the Brain of Rats with TBI that are Ubiquinol-Treated Compared to Untreated.

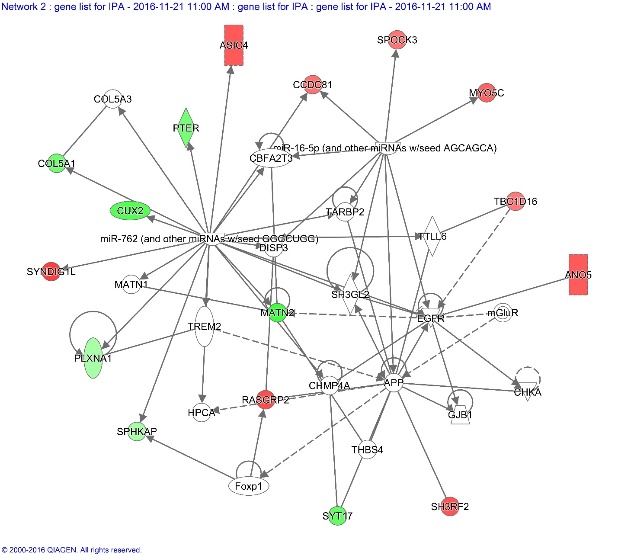
**Regulatory effects**

The regulator effects function links the upstream regulators to dataset molecules to downstream diseases and functions to model pathway interactions [12]. This function helps identify specific hypotheses by creating a depiction of the biological mechanisms and identifying the resulting clinical phenotype. The clinician completing the analysis can customize the regulator effects network by selecting specific regulators and diseases or functions. The relationship between activated or inhibited upstream regulators with the impact on downstream biology can result in discovery of new regulators and confirm known regulators with clinical endpoints [5].

**Networks**

A network is a set of genes with relationships that are generated based on input data [12]. Networks are developed from focus molecules that have the most interactions with other focus molecules [21]. Creating networks identifies which molecules have the most relationships and therefore could be the most biologically relevant [22]. Each network has an identifying network score based on the number of focus molecules. The higher the score, the less likely the observed molecules from the network appeared in the data by chance [16,21]. Overlapping networks can be used to identifying molecules that are present in multiple networks and a single network is developed from merged networks [5].

In the TBI study, 67 transcripts were differentiated in the ubiquinol group. Of these transcripts, 26 were upregulated and 41 were down regulated. (Figure 3) is an example network with miR-762 as the focus molecule. A total of nine genes were upregulated (SYNDIG1L, ASIC4, CCDC81, SPOCK3, MYO5C, TBC1D16, ANO5, SH3RF2, and RASGRP2) and seven genes were downregulated (COL5A1, PTER, CUX2, PLXNA1, MATN2, SPHKAP, and SYT17) [9]. MiR-762 was observed to regulate mitochondrial function, adenosine triphosphate levels, and apoptosis in the heart [23]. While the role of miRNAs in the nervous system remains unclear, they are implicated in the translation, RNA metabolism, gene development, and regulation of nervous system functionality [24].



**Figure 3:** Example Network of Differentially Expressed Genes in the Brain of Rats with TBI that are Ubiquinol-Treated Compared to Untreated. Red-labeled boxes: upregulated genes. Green-labeled: downregulated genes.

**Conclusion**

IPA is a powerful web-based analytical software application to integrate and analyze massive amount of omics data. In contrast with many existing software designed for pathways analysis, IPA utilizes the quantitative information generated from multi-omics experiments to produce meaningful biological results. As illustrated in this article, the five core analysis components of IPA provide comprehensive and thorough bioinformatic analyses of omics data. This helps to interpret and understand the significance of the data and identify genes, molecules or biomarkers underlying disease mechanisms. With improved understanding of IPA, clinicians are better prepared to lead innovative clinical efforts to advance precision health, aiming to optimize patients’ health outcomes.

**References**

1. [Fu MR, Kurnat-Thoma E, Starkweather A, et al. (2020) Precision health: A nursing perspective 7: 5-12.](https://pubmed.ncbi.nlm.nih.gov/32099853/)
2. [Jensen MA, Ferretti V, Grossman RL, Staudt LM, et al. (2014) The NCI Genomic Data Commons as an engine for precision medicine Blood 130: 453-9.](https://pubmed.ncbi.nlm.nih.gov/28600341/)
3. [Sandhu C, Qureshi A, Emili A (2018) Panomics for Precision Medicine. Trends Mol Med 24: 85-101.](https://pubmed.ncbi.nlm.nih.gov/29217119/)
4. [Thomas S, Bonchev D (2010) A survey of current software for network analysis in molecular biology. Hum Genomics 4: 353-60.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3500165/)
5. [Qiagen. IPA User Manual.](https://qiagen.secure.force.com/KnowledgeBase/KnowledgeIPAPage)
6. [Bubier JA, Sutphin GL, Reynolds TJ, et al. (2019) Integration of heterogeneous functional genomics data in gerontology research to find genes and pathway underlying aging across species. PLoS One 14: e0214523.](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214523)
7. [Cirillo E, Parnell LD, Evelo CT (2017) A Review of Pathway-Based Analysis Tools That Visualize Genetic Variants. Front Genet 8:174.](https://www.frontiersin.org/articles/10.3389/fgene.2017.00174/full)
8. [Kramer A, Green J, Pollard J, et al. (2014) Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30: 523-30.](https://academic.oup.com/bioinformatics/article/30/4/523/202720)
9. [Pierce JD, Shen Q, Peltzer J, et al. (2017) A pilot study exploring the effects of ubiquinol on brain genomics after traumatic brain injury. Nurs Outlook 65: S44-S52.](https://pubmed.ncbi.nlm.nih.gov/28755974/)
10. [Ackers I, Malgor R (2018) Interrelationship of canonical and non-canonical Wnt signalling pathways in chronic metabolic diseases. Diab Vasc Dis Res 15: 3-13.](https://pubmed.ncbi.nlm.nih.gov/29113510/)
11. [Shao Z, Wang K, Zhang S , et al. (2020) Ingenuity pathway analysis of differentially expressed genes involved in signaling pathways and molecular networks in RhoE geneedited cardiomyocytes. Int J Mol Med 46:1225-38.](https://www.spandidos-publications.com/10.3892/ijmm.2020.4661)
12. [Wu C, Chen M, Zhang Q, et al. (2019) Genomic and GeneChip expression profiling reveals the inhibitory effects of Amorphophalli Rhizoma in TNBC cells. J Ethnopharmacol 235: 206-18.](https://pubmed.ncbi.nlm.nih.gov/30853353/)
13. [Farashi S, Kryza T, Batra J (2020) Pathway Analysis of Genes Identified through Post-GWAS to Underpin Prostate Cancer Aetiology. Genes (Basel) 11.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7291227/)
14. [Li X, Long J, He T, et al. (2017) Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease. Sci Rep 5:12393.](https://www.nature.com/articles/srep12393)
15. [Valkovic AL, Bathgate RA, Samuel CS, et al. (2019) Understanding relaxin signalling at the cellular level. Mol Cell Endocrinol 487: 24-33.](https://pubmed.ncbi.nlm.nih.gov/30592984/)
16. [Vishnubalaji R, Shaath H, Alajez NM (2020) Protein Coding and Long Noncoding RNA (lncRNA) Transcriptional Landscape in SARS-CoV-2 Infected Bronchial Epithelial Cells Highlight a Role for Interferon and Inflammatory Response. Genes (Basel) 11.](https://pubmed.ncbi.nlm.nih.gov/32646047/)
17. [Yu J, Gu X, Yi S (2016) Ingenuity Pathway Analysis of Gene Expression Profiles in Distal Nerve Stump following Nerve Injury: Insights into Wallerian Degeneration. Front Cell Neurosci 10: 274.](https://www.frontiersin.org/articles/10.3389/fncel.2016.00274/full)
18. [Khan MI, Debski KJ, Dabrowski M, et al. (2016) Gene set enrichment analysis and ingenuity pathway analysis of metastatic clear cell renal cell carcinoma cell line. Am J Physiol Renal Physiol 311: F424-36.](https://pubmed.ncbi.nlm.nih.gov/27279483/)
19. [Li JY, Yong YH, Gong DL, et al. (2019) Proteomic analysis of the response of porcine adrenal gland to heat stress. Res Vet Sci 122: 102-10.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7111741/)
20. [Pang K, Hao L, Shi Z, et al. (2020) Comprehensive gene expression analysis after ERH gene knockdown in human bladder cancer T24 cell lines. Gene 738: 144475.](https://pubmed.ncbi.nlm.nih.gov/32081697/)
21. [Jovov B, Araujo-Perez F, Sigel CS, et al. (2012) Differential gene expression between African American and European American colorectal cancer patients. PLoS One 7: e30168.](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0030168)
22. [Ramanan VK, Shen L, Moore JH, et al. (2012) Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet 28: 323-32.](https://pubmed.ncbi.nlm.nih.gov/22480918/)
23. [Yan K, An T, Zhai M, et al. (2019) Mitochondrial miR-762 regulates apoptosis and myocardial infarction by impairing ND2. Cell Death Dis 10: 500.](https://www.nature.com/articles/s41419-019-1734-7)
24. [Condrat CE, Thompson DC, Barbu MG, et al. (2020) miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells 9.](https://www.thermofisher.com/procalcitonin/wo/en/home.html?cid=0se_gaw_25052021_DA7MGR)