# Current Trends and Terminology in Systems Biology and Pathway Research:

# A Conceptual Roadmap

By

Michael J. Ames, BS

**Capstone Project**

Submitted in partial fulfillment of the requirements for the degree of

Master of Biomedical Informatics

Oregon Health and Science University

May 2010

School of Medicine

Oregon Health & Science University

---

CERTIFICATE OF APPROVAL

---

This is to certify that the Master's Capstone Project of

Michael J. Ames

*"Current Trends and Terminology in Systems Biology and Pathway Research: A Conceptual Roadmap"*

Has been approved

Shannon McWeeney, PhD
Capstone Advisor

# TABLE OF CONTENTS

# LIST OF TABLES AND FIGURES

## Tables



## Figures

## ACKNOWLEDGEMENTS

Everyone knows Rule #1 when selecting a topic for your capstone project: Make sure the subject matter is something you already know. Research and writing is difficult enough without having to learn an entire new subject. But in a fit of scientific enthusiasm, I violated Rule #1 out of the gate and selected for my topic something that fascinated and inspired me but about which I actually knew very little. It is for this reason that I express deepest gratitude to my capstone advisor, Dr. Shannon McWeeney. For over a year, she patiently guided my research ("You know what proteins are, right?") and resolved my misunderstandings ("It's called the Fisher *Exact* Test, Michael. *Exact!*") She did all this across the telephone and Internet as I juggled a complicated work and family schedule, a career change, and relocation to a new state. Shannon, thank you.

I also wish to acknowledge the encouragement and support of Jessica Bondy, my supervisor at the University of Colorado Cancer Center, as well as my fellow team members in the Research Informatics Core. Thank you for putting up with unprecedented (and hopefully unrepeated) levels of distractedness on my part, especially during the last few weeks of preparing this project. The work we do together inspires me beyond words.

Finally, I wish to acknowledge my wife, Libby, who succeeded where all others had failed in teaching me how to be a student. She has paid for it dearly as we have spent almost every day of our married life with my next homework assignment, next test, next class, next project, lurking around the next corner. Here's to a moment's respite. Your turn next.

## ABSTRACT

**Introduction**. The past decade has witnessed exceptional growth in research involving the study of biological pathways and corresponding growth in bioinformatics tools to support it. Evolving concepts in the field are not always understood consistently, increasing the risk of misinterpreted or lost research. It is therefore imperative to bioinformaticians that possible areas of confusion be identified and clarified.

**Background**. Pathway research defines or refines our understanding of interactions between molecules within and between cells, primarily to identify molecular targets for disease therapies. Pathways interact to form networks. Pathways and networks should be understood as concrete series of chemical reactions, but as abstract, probabilistic models of molecular behavior.

**Methods**. Three top peer-reviewed bioinformatics journals were selected. Articles published over a one year period of time were reviewed to identify the focus of the article, the type of research being conducted, and specific terminology used.

**Results and Discussion**. The most common research activities were predictive modeling, curation and pathway enrichment analysis. Predictive modeling is distinguished from descriptive modeling by its exploratory nature. The objective of curation is to integrate disparate data sources for the purpose of reuse. Enrichment analysis can be performed with either pathway or functional annotations, but care must be taken not to confuse the two.

**Conclusion**. Pathway research consists of a variety of activities and will continue to evolve. With clear understanding of these research concepts, bioinformaticians can more effectively communicate and disseminate tools to biologists to help realize the full potential of pathway research and systems biology.

## INTRODUCTION

Researchers in the field of Systems Biology study the biological mechanisms underlying human health across various scales and levels. At the most granular level, biological processes can be understood as molecular interactions within and between cells. Generally speaking, a known sequence of molecular interactions is called a pathway. Disruption of path is believed to be the cause of a wide range of diseases, so pathways are a key unit of study in the search for more effective therapies.

Recent years have witnessed an explosion in technology used to infer the presence and behavior of genes and proteins within the body, such as gene expression microarrays, next-generation DNA and RNA sequencing systems, and proteomics. The vast data generated by these systems are driving a corresponding increase in the development of bioinformatics systems used to store and analyze it. This paper seeks to aid bioinformaticians in developing a clear understanding of current trends in systems biology and pathway research, the major types of research being conducted, and key terminology associated with it.

To estimate both the significance and emerging nature of systems biology and pathway research, counts of relevant journal articles in PubMed were conducted. First, PubMed was queried for journal articles containing terms commonly associated with pathway research from the years 2000 to 2009 (see Appendix A). As illustrated in Figure 1, these publications have increased at an average rate of approximately 12% each year, more than doubling from 4,998 to 13,489.

Next, to approximate the increasing usage of bioinformatics tools in supporting pathway research, the initial set of articles were limited further by the presence of terms commonly

associated with bioinformatics tools (see Appendix B). These numbers are even more compelling, increasing by nearly 40% each year from 61 articles in 2000 to 1061 articles in 2009, as illustrated in Figure 2.

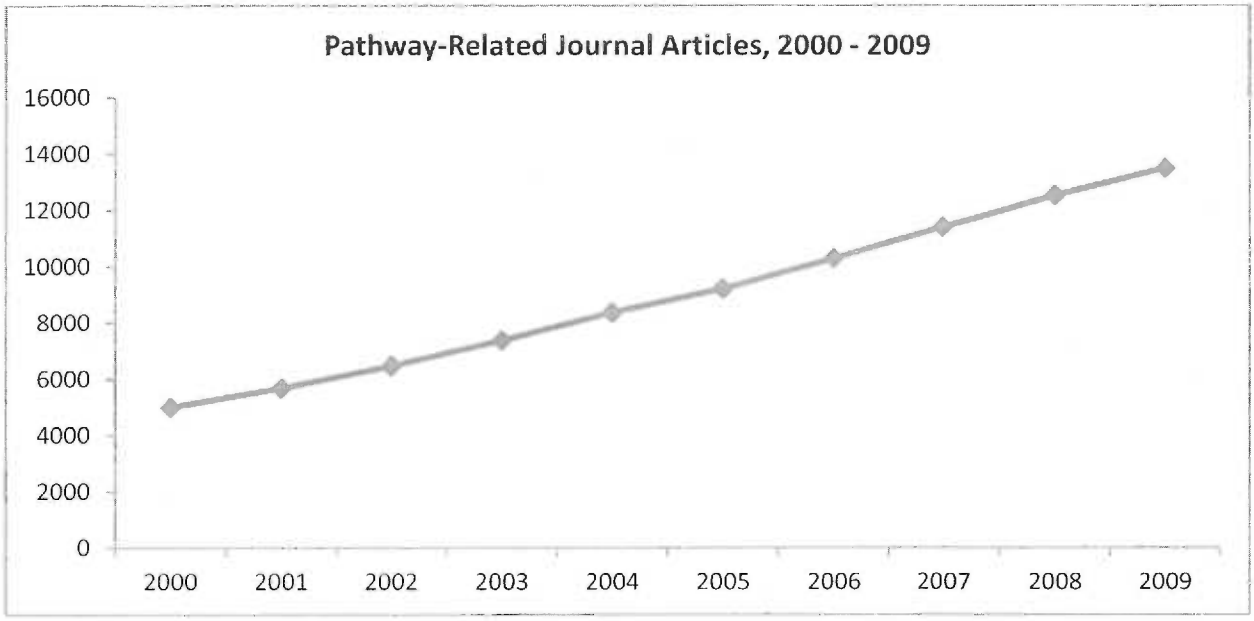


**Figure 1: Counts of Pathway-Related Journal Articles**

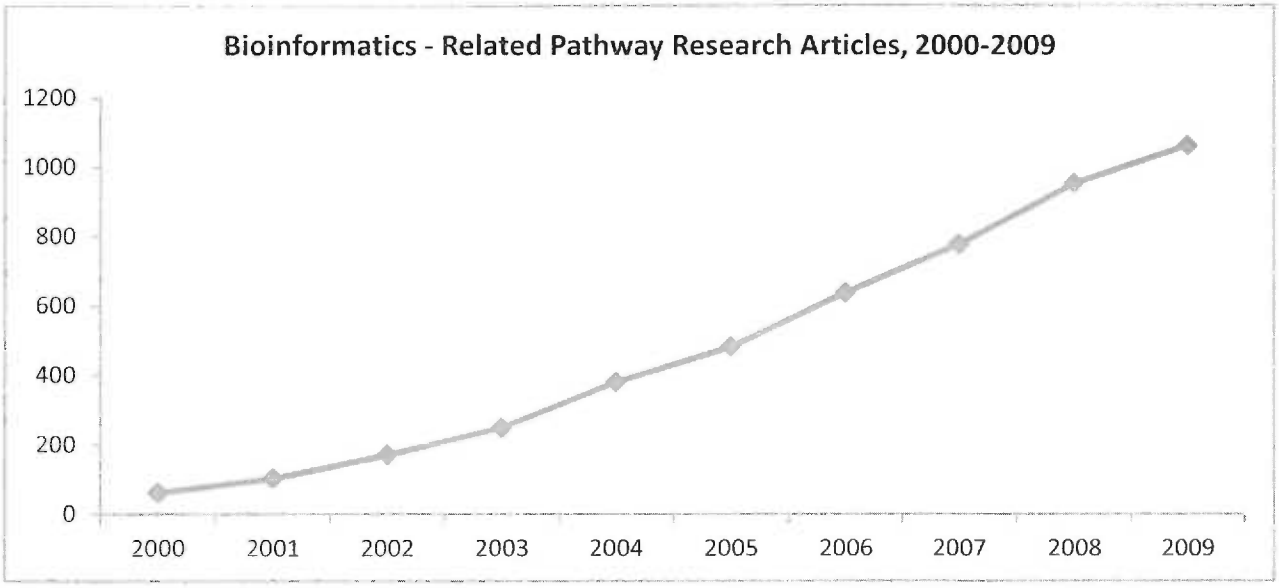


**Figure 2: Counts of Bioinformatics-Related Pathway Research Articles**

Finally, the ratio of the two article counts year over year was taken to show the percentage of pathway research overall specifically citing bioinformatics tools. In 2000, activities of this type accounted for approximately 1.2% of published pathway research articles. As illustrated in Figure 3, by 2009 these activities increased by 650% accounted for nearly 7.9% -- a 6.5-fold increase.

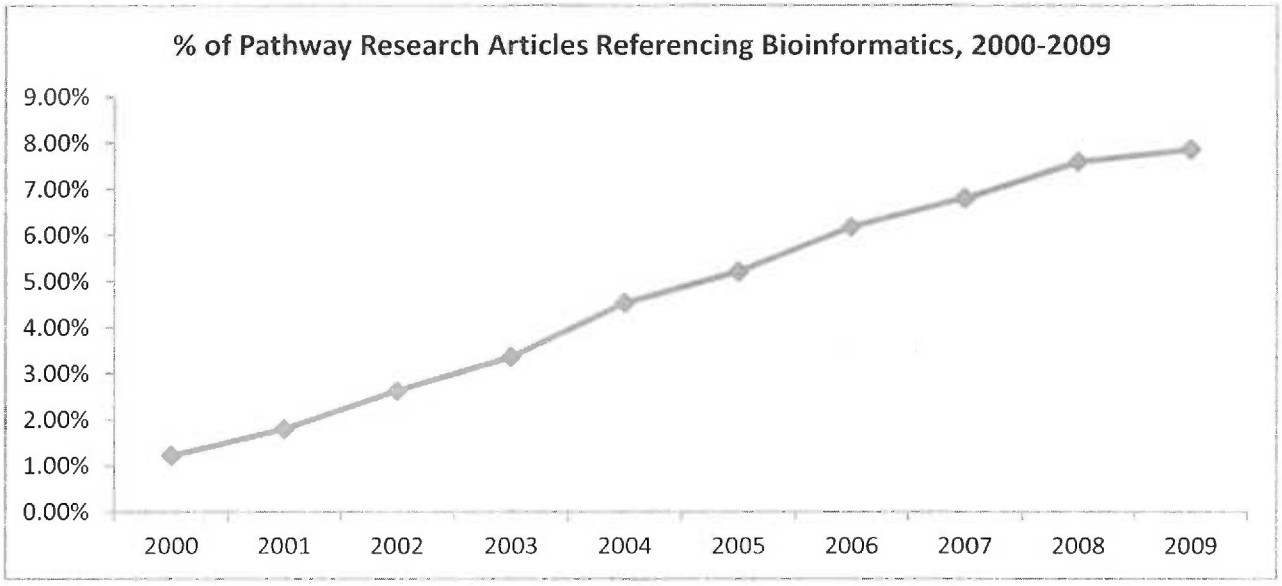


**Figure 3: Percent of Pathway-Related Articles Incorporating Bioinformatics Terms**

With growth comes change. The techniques, tools and terminology used by researchers are rapidly evolving. This is inevitable in any emerging science, but presents certain risks to researchers. At best, such ambiguities introduce confusion and inefficiency into the research process. At worst, they lead to incorrect interpretation of findings and, ultimately, loss of research. Therefore, it is imperative to bioinformaticians that possible areas of confusion be identified as early as possible so that research activities and outcomes can be communicated unambiguously.

## BACKGROUND

For centuries, medical research was based exclusively on human-observable disease characteristics.[1] This improved over time as laboratory analyses were developed to measure characteristics beyond the direct reach of human senses. Hypotheses were rigorously tested and refined, resulting in remarkable medical breakthroughs, such as the curing of infections with antibiotics and the virtual elimination of poliomyelitis. Interestingly, this "top-down" approach to medical research tended to lead researchers to understand *what* treatments worked before understanding *how* they worked. The chemical structure of penicillin, for example, was not described until nearly a decade after it was first used in medical practice.[2]

Increased understanding of the role of molecular processes in disease led to the field of Systems Biology in the mid-1900s. The premise of systems biology is that biological processes can be understood better with a "bottom-up" approach; i.e., start by understanding the basic interplay of molecules at the cellular level, then form building blocks of systems that can be used to describe increasingly large and complex processes.[3] In the context of medical research, this means first seeking to understand *how* a disease process works and then identifying *what* treatments might be effective in remediating the errant process.

The potential for this approach to yield meaningful treatments was dramatically accelerated later in the century with the advent of gene microarrays and other "high-throughput" systems. These laboratory systems analyze biological material and generate massive amounts of data that can be used to help form computerized models of molecular systems, leading to greatly increased possibilities for identifying target molecules for drug therapies.

Initially, research centered on attempting to correlate disease phenotypes with the presence – or lack of presence – of individual genes or proteins. However, as more came to be understood about the nature of molecular interactions in the body, this single-gene approach to understanding disease proved to be limited in its ability to find meaningful drug targets. Disease is often – if not usually – the result of a combination of genetic factors; i.e, multiple genes and proteins interact together to produce a given phenotypic state. Understanding the activities of a single gene is not enough to predict or alter outcomes in a diseased cell. Rather, researchers must understand the activities of that gene in the context of its molecular *system* – the complete set of interactions between that gene and other molecules it comes in contact with.

The study of these systems has yielded some compelling results in recent years. A few examples are:

- An experiment in which pathways were analyzed in the identification of the androgen receptor gene (AR) as both a mediator of prostate cancer as well as an indicator of the cancer's level of aggressiveness.[4]
- Genome-wide association studies (GWAS) in which pathway enrichment analysis was used to prioritize long lists of disease-implicated genes, enabling researchers to focus on those genes most likely to yield meaningful therapeutic targets.[5]
- A study was conducted examining known pathways associated with coronary arteriosclerosis. This study identified the gene interleukin 6 (IL6) as a strong candidate for a therapy to treat arteriosclerosis in diabetic patients.[6]

These examples illustrate the promise of pathway research. To fulfill this promise, further development of computational tools and techniques needed is needed to keep pace with evolving technologies and data types.

## METHODS

A literature review and analysis was conducted to quantify major trends and terminology in molecular systems research. To ensure a cohort of articles both sensitive enough cover a wide range of research activities and specific enough to reflect major trends, articles were selected meeting all of the following criteria (see Appendix C):

- Indexed in PubMed
- Peer-reviewed journal articles
- Articles from the three journals with the highest impact factor per the ISI Web of Knowledge.[7]
- Articles published in a recent one-year period.
- Articles referencing the term "pathway" in any field

The abstract and full text for each article were read and categorized according to each of the following attributes:

- Article relevance. One of two options:
    - Yes: The article references molecular pathways in some fashion.
    - No: The article does not reference molecular pathways.
- Focus of article. One of two options:
    - Method or study: The focus of the article is to describe a study or experiment that was performed or a method or process for studies and experiments.
    - Tool or resource: The focus of the article is to describe a bioinformatics tool or resource that has been developed.

- Terminology usage. Zero, one or both of the following:
    - The article incorporates the term "pathway analysis" in any field.
    - The article incorporates the term "network" in the sense of molecular interactions, in any field.
- Primary research activities. One or more of the following classifications, which were developed from an initial survey of a subset of articles (see Results and Discussion for further details):
    - Descriptive modeling
    - Predictive modeling
    - Curation
    - Pathway enrichment
    - Functional enrichment
    - Other

The results of this literature review serve as the quantitative basis for the concepts and issues discussed in the remainder of this paper.

## RESULTS AND DISCUSSION

### Describing Molecular Systems: Pathways and Networks

Of the 100 papers reviewed, all of which used the term "pathway" in some way, 53 also used the term "network" to imply similar meaning. The question presents itself: Are pathways and networks synonyms for the same concept, mutually exclusive concepts, or somehow interrelated? Of the articles reviewed, examples existed for all three possibilities – clearly concepts in need of clarification.

Researchers generally define pathways and networks as separate, but related concepts.[8] One article defined pathways as a series of "consecutive reaction steps" but distinguishes them from networks in that pathways can be defined and studied in different contexts (cell type, tissue type, etc.) while networks cannot. According to this definition, a pathway is a definable, reusable sequence of steps that one might expect to result in consistent outcomes across experiments. A network, in contrast, is "built *de novo*" with each experiment and can only be understood in its specific context (time, tissue type, cell type, disease state, etc.)

These definitions need clarification in at least two ways: First, by emphasizing that networks are subject to contextual changes, some may incorrectly conclude that pathways are not (see *Modeling* for further discussion on this point). Second, networks are not all strictly *de novo* constructs of single experiments. While networks, like pathways, must always be interpreted in light of the context in which they are examined, networks may have elements that can be consistently predicted and replicated across a variety of contexts.

Another challenge in defining pathways arises in defining the scope, or boundaries of individual pathways. Are the sets of chemical reactions that comprise pathways actual biological

entities with self-defining boundaries, in the same sense that one bodily organ can be distinguished from another? Are they the result of limitations in what can be observed and measured? A researcher's arbitrary point of view? This issue can have significant impact on research outcomes. Various pathway databases define pathway boundaries differently, and at least one study has concluded that outcomes of pathway analyses can differ dramatically depending on which database – and, therefore, which approach to boundaries – is used.[9]

One of the more well-considered attempts at defining pathways and networks was the result of the efforts of over 30 participants with various backgrounds in biomedicine and bioinformatics.[10] This group considered several competing definitions of pathway, and offered the following definition: "A connected sequence of two or more processes having a shared causality in that the processes contribute to realizing a common function, whereby the output of one process is the input for the next process in the sequence."

Parsing this definition yields a few key attributes of a pathway:

- *Sequence*. A list of genes known or supposed to interact is *not* pathway unless it incorporates some concept of the sequence of those interactions.
- *Common causality*. One way a pathway boundary is determined is by the causes that trigger it – a single pathway is the result of shared causal events.
- *Common function*. Pathways are *not* self-defined entities. Rather, they are rather defined for purposes of research based on the fact that the interactions involved work to bring about a common function.

Similarly, a network is defined as "a connected sequence of two or more pathways having a shared causality in that the pathways contribute to realizing a common function, and

involving distinct pathways, where the output of one pathway is the input for the next pathway in the sequence." In other words, networks are comprised of pathways and bounded on the same criteria – common causality and function.

One example of effective use of these definitions was a paper comparing various algorithmic approaches for identifying interacting pathways with a network.[11] Another applied related algorithms to identify individual pathways from within a known set of protein-protein interactions (PPIs) affecting angiogenesis.[12]

In short, both pathways and networks are sequenced series' of interactions bounded by common causality and function. Pathways are comprised of interactions between molecules, and networks are comprised of interactions between pathways.

**Interpreting Pathways: Concrete Processes vs. Abstract Models**

Pathway models are used to simulate and communicate the interactions that occur within a pathway. These models might be in machine readable form, such as GenMAPP Pathway Markup Language (GPML)[13] or take the form of images, such as the Signal Transduction Pathways diagram shown in Figure 4. These concise diagrams are an indispensable means of documenting knowledge concerning pathway members and their sequence of interactions. In a glance, one can see the pathway boundaries, participating genes, and interactions with other pathways.

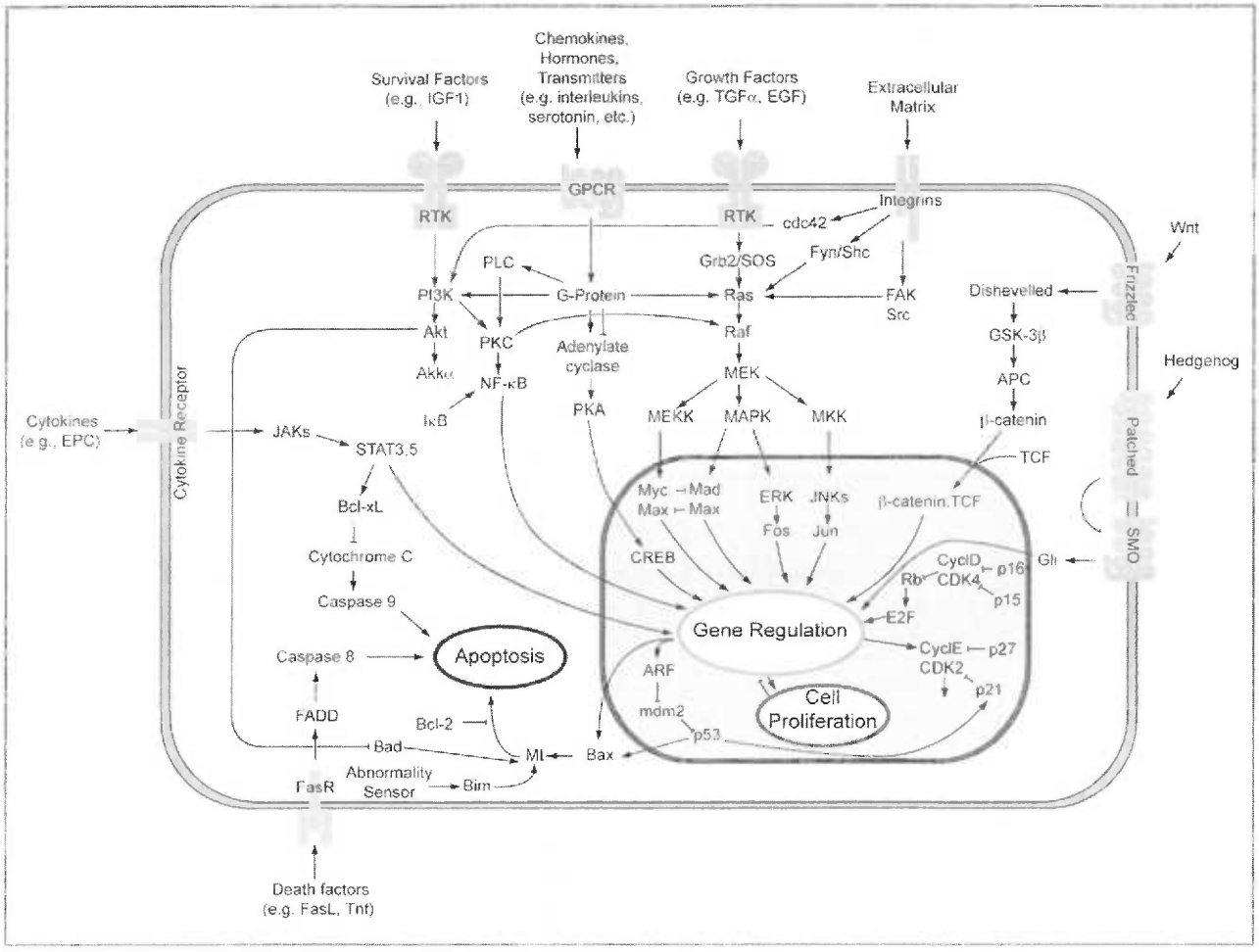


**Figure 4: Visual Model of Signal Transduction Pathways**[14]

However, users of such diagrams may not be aware of the abstract, context-sensitive nature of all pathway models. Figure 4, for example, shows a direct interaction between "JAKs" and "STAT3,5." In reality, this pathway consists of multiple, differing interactions between at least four separate molecules in the JAK and STAT families, but for simplicity in the figure, on two nodes are represented. Furthermore, research has shown that these interactions are modulated by another gene, BRCA1, which appears nowhere on the diagram.[15]

The missing information on this diagram is by no means an error or oversight on the author's part. Indeed, one of the first decisions a researcher must make in developing a pathway model is the appropriate level of detail to include – a balance between providing a meaningful basis for further research vs. the complexity involved in developing a more detailed model.[16] Modeling all aspects of every interaction in a pathway would be a monumental task. As one author put it, the underlying "biological truth" of a pathway "is so complex that all models are, in a strict sense, wrong."[17]

In addition to being intentionally simplified, pathway models should not be interpreted as concrete processes – sequences of well-defined, highly reproducible chemical reactions. Rather, they are probabilistic models highly subject to the context in which they take place. Species, cell type, disease state, environmental conditions, and a plethora of other factors can significantly alter the operation of a defined pathway. Also, errors and inconsistencies may exist in the source data used in generating the model, resulting in nodes and regions with greater and lesser degrees of confidence. This level of detail is often omitted from the visualization for the sake of simplicity.

Failing to consider the abstract nature of a pathway model could lead a researcher to false conclusions if, for example, a laboratory experiment fails to produce results consistent with a published pathway. Concluding that the published pathway is inaccurate would be premature – it is possible that the pathway was simply operating in an altered state due to any number of contextual factors. Bioinformaticians should carefully consider the abstract, probabilistic nature of these models as they analyze and develop tools to study them.

## Pathway Research: A Variety of Activities

The literature review revealed a variety of pathway research types. This diversity of activities is consistent with a cyclical model of systems biology research proposed in 2002. In it, "wet" laboratory experiments lead to data and hypotheses that are modeled and simulated in "dry" computational experiments. These experiments lead to predictions which are confirmed and refined in further laboratory experiments. In other words, systems biology requires a high degree of both laboratory and *in silico* experimentation, with many different techniques and disciplines along the way. Some examples are:

- A paper was published describing a new computational method for inferring pathways from large amounts of genomic data.[18]
- An ovarian cancer study used a database of pathways to enrich results from a gene expression study.[19]
- An automated text-mining approach was used to extend pathway membership from molecular interactions published in journals.[20]

However, to accurately convey the complete pathway research landscape, some additional concepts must be incorporated into the 2002 model. Specifically, the literature review revealed a sizeable effort around pathway curation (defined in more detail below), resulting in ever-growing resources of reusable pathway information. While this is a component of the "dry" computational experiments described above, it occurs frequently enough in pathway research that it warrants more emphasis. Also, while "wet" laboratory experiments remain fundamental to testing and proving hypotheses in the lab, they are only one part of the research life cycle to bioinformaticians.

Figure 5 presents a modified view of the cyclical systems biology research cycle with the aforementioned modifications in place. This cycle is highly dynamic –experimental data can flow directly into any of the three major research activities, and results of those activities can directly influence further experiments. Also note that results from curation, modeling, and analysis are all candidates for publication to literature and/or pathway databases.

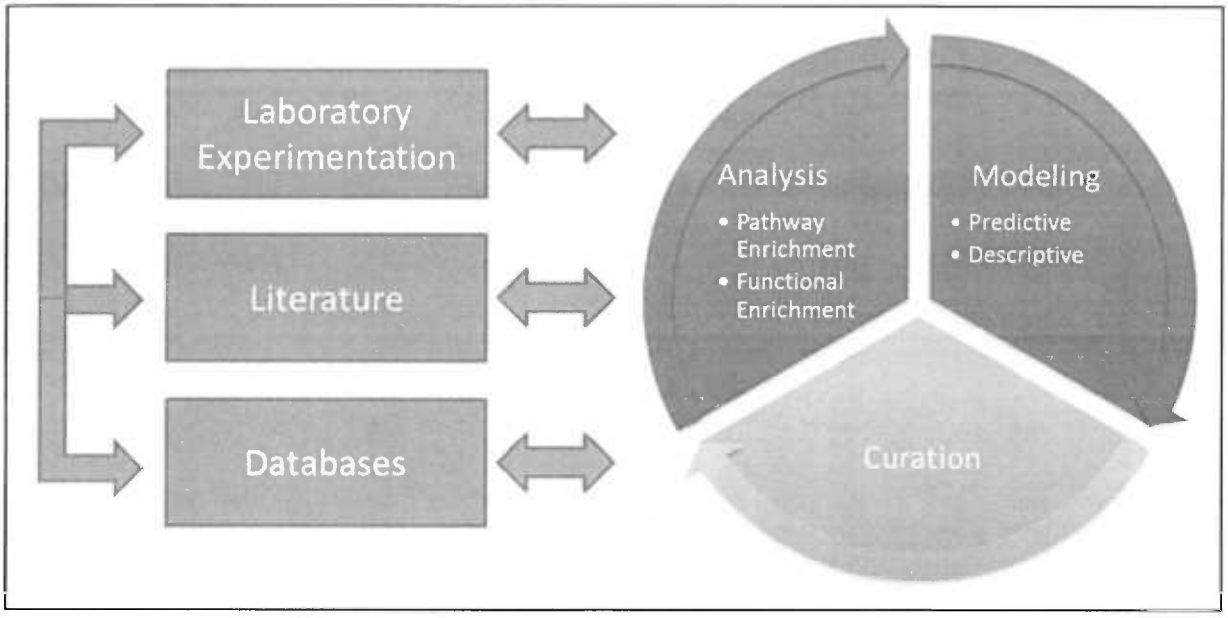


**Figure 5: Overview of the Pathway Research Cycle**

Figure 6 shows the frequency at which various pathway research activities were the primary topic of each articles in the literature review. Predictive modeling, curation and pathway enrichment were the most frequently described activities. The total count (103) is slightly higher than the total number of articles reviewed (100) because some articles encompassed more than one significant activity. Discussion and definitions for each of these activity types will be given in the following sections.

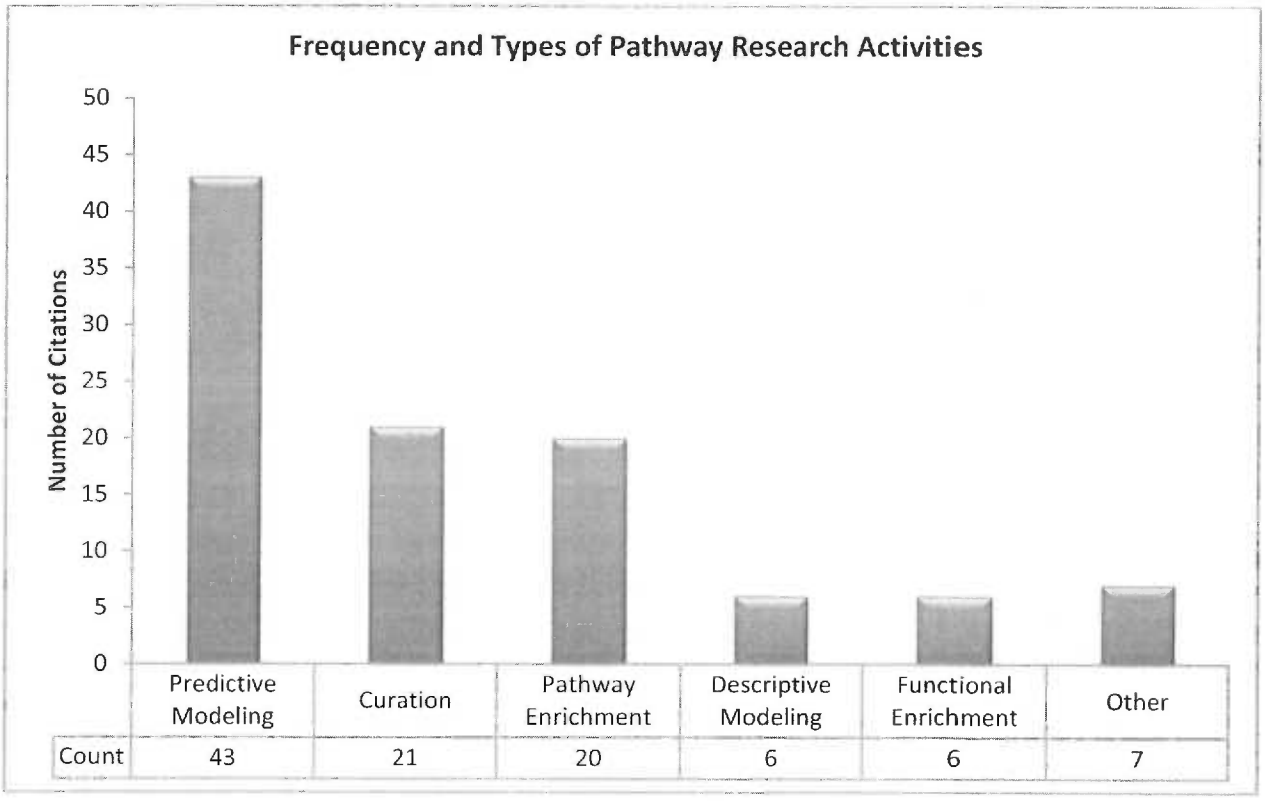


**Figure 6: Frequency of Types of Research Activities**

## Modeling

Predictive modeling is the use of computational techniques to analyze data for the purpose of discovering, extending or simulating pathways and networks. The literature review revealed predictive modeling as the most common type of pathway research activity, with 43 of 100 articles focusing on predictive modeling as the primary topic. Of these, the majority (30) were study or method papers. This is consistent with the exploratory nature of predictive modeling – techniques vary widely, and may be difficult to encapsulate in reusable software. Some examples of predictive modeling follow:

1. A paper was published describing a method for discovering pathways using a computational approach called subgraph extraction.[11] This technique involves integrating large amounts of source experimental data and processing it with a complex algorithm for weighting and predicting connections between pathway nodes. Inferred pathways produced by variations on this algorithm were compared to a reference pathway in order to measure the utility of the approach.
2. A technique was described for identifying "characteristic sub pathway networks," a novel concept used in helping to surface interactions between pathways based on protein-protein interactions.[12] While this study incorporated pathway enrichment (described below), it went well beyond enrichment in applying computationally intensive algorithms to elucidate interaction points between pathways. This information was used to construct a comprehensive network of pathways involved in the promotion of angiogenesis.
3. A Bayesian learning algorithm was used in conjunction with data from gene knockdown experiments using RNAi screening technology.[21] This technique was used to predict the placement of known pathway members within the sequence of interactions forming the pathway.

As seen by these examples, the specific experiments, data, statistics and algorithms employed can vary widely from one modeling project to the next. These projects often incorporate multidisciplinary teams of researchers bringing together data from a variety of experiments and diverse expertise in molecular biology, bioinformatics, statistics and computer science. But the common thread in these activities is the use of data-rich and computationally intensive techniques in order to predict *previously unknown* information about pathways.

Contrast this with another use of the term "modeling" also found in the literature – the documentation of *known* pathways in machine-readable or visual form. This form of modeling was the focus of six articles; however, many pathway-related articles incorporate visual models of pathways within the article, so it is clearly a common activity. It differs from the computational modeling described above in that its purpose is descriptive, rather than predictive. Predictive modeling seeks to discover new information about pathways, while descriptive modeling seeks to document and communicate that information.

Many software systems exist to aid in descriptive modeling and visualization. In fact, all six descriptive modeling articles were specifically describing a tool or resource designed to aid in the process. Some of these tools include GenMAPP,[13] Cytoscape,[22] and ChiBE.[23] The capabilities of these systems vary widely, and many include features which can be used to aid in predictive modeling as well.[24]

Distinguishing between the two types of modeling is important in identifying appropriate tools for the question of interest. Predictive modeling requires computational tools to support inference, exploration and simulation of pathways. Most often, descriptive modeling requires visualization tools to create pathway and network diagrams. A relationship exists – predictive modeling tools often export information in a format that can be used with descriptive modeling tools; however, while visually documenting a pathway is a likely outcome of predictive modeling, it is by no means required.

The following definitions will aid bioinformaticians in clearly communicating the intent of modeling activities:

- *Predictive Modeling*: Discovery, refinement, extension or simulation pathways through computational tools.
- *Descriptive Modeling*: Documentation of known pathways in machine-readable or visual form for the purpose of communication, visualization, curation and/or publication.

**Curation**

The next most common research activity discussed in the literature is curation, with 21 of 100 articles referencing it as a primary research activity. As illustrated in Figure 5, curation is both an input to, and a result of, modeling and analysis activities. Curation can take many forms, as illustrated in the following examples:

1. A publication described a software system developed to aid researchers in searching published literature for information relevant to yeast metabolic pathways so that it could be incorporated into descriptive models. The system did not automatically extract data, but provided a means for more efficiently conducting manual searches to aid in fully documenting pathways.[25]
2. A detailed workflow and natural language processing algorithm were implemented to mine protein and molecular interactions from published literature on carotenoid and Vitamin A metabolism. The mined results were reviewed by domain experts and used to expand and clarify existing models of these pathways.[26]
3. A web-based application was developed to allow researchers to collaborate in documenting pathways. The application allows users to describe pathways in visual

models as wells as structured data formats. Further, pathways can be annotated with additional arbitrary experimental data.[27]

As illustrated by these examples, inputs to the curation process may include the mental knowledge of domain experts, research results written in natural language, vast quantities of experimental data or other heterogeneous sources of information. As its output, curation distills this data into structured data and/or visual models that can be efficiently reused in further research. Typically, these models are incorporated into a central database, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG)[28] or Reactome.[29]

The act of increasing the value of knowledge through integrating disparate data sources in a reusable way is key to distinguishing curation from other activities.[30] A predictive modeling simulation can make use of a curated pathway model as an input; however, unless the knowledge gained from the simulation is incorporated into a structured, reusable format, with potential inconsistencies reconciled, it has not been curated.

Conversely, research activities that consist solely of integrating existing knowledge are most likely curation activities and should not be described as either modeling or analysis. Failing to make this distinction can create confusion about the intended purpose of the activity. Predictive modeling and pathway analysis result in new information about pathways. Curation distills that knowledge into an unambiguous, reusable form to aid in further research.

**Analysis**

The final common form of pathway research identified by the literature review is most commonly referred to as "pathway analysis." This term, however, is problematic, as it is also frequently used to describe predictive modeling and even some forms of non-pathway-related research. So to be more specific, some form of *enrichment analysis* was cited as a primary research type in 23 of 100 distinct articles.

The goal of enrichment analysis is determine if a pathway is implicated in the presence of a known phenotype or to determine if a gene may be a member of a known pathway. In its most common form, enrichment analysis is carried out as follows:

1. Gene microarray experiments are conducted to develop lists of genes expressed differently in control and test cells.
2. Differentially expressed genes are annotated with their membership in known pathways.
3. The significance of the differential expression levels is calculated using various computational methods, such as Fisher's Exact Test or Gene Set Enrichment Analysis (GSEA).[31 32]

The second step, annotation, is usually accomplished with pathway membership as contained in a pathway database such as KEGG. Annotation with pathway sources occurred in 20 of the 26 enrichment-related articles reviewed. However, in 6 articles, annotation was performed using biological functions or processes from the Gene Ontology (GO). A given study may perform enrichment analysis using pathway membership, biological functions, or both.

In some instances, enrichment analyses were described as pathway analyses yet only included annotations from GO.[33] [34] The rationale behind this is unclear, but it is possible that the distinction between pathways and biological functions is confusing with respect to enrichment analysis. According to GO's documentation, "A biological process is not equivalent to a pathway; at present, GO does not try to represent the dynamics or dependencies that would be required to fully describe a pathway."[35] It has been suggested that issues such as this may arise from a lack of understanding of ontologies in general as well as weaknesses in how biological ontologies – even successful ones, such as GO – have been organized.[36]

One way to illustrate the difference between pathway and functional databases is to search for a particular gene in both types of database and compare the results. For example, searching for PLCB1 (phospholipase C, beta 1) in KEGG returns its pathway membership as shown in Table 1. Contrast this with the list of functional annotations returned when searching for the same gene in the GO, as shown in Table 2. While a few items from both lists seem to be related; e.g., Alzheimer's disease (KEGG) and memory (GO), the content of the two lists is significantly different. This is because pathways are related to biological processes and functions, but are not the same concepts. Pathways are some of the building blocks that lead to biological processes and functions, but implication of a gene in a particular biological function does not imply membership in any particular pathway.

| Pathway | |
|---|---|
| Inositol phosphate metabolism | Gap junction |
| Metabolic pathways | Long-term potentiation |
| Calcium signaling pathway | Long-term depression |
| Chemokine signaling pathway | GnRH signaling pathway |
| Phosphatidylinositol signaling system | Melanogenesis |
| Vascular smooth muscle contraction | Alzheimer's disease |
| Wnt signaling pathway | Huntington's disease |

**Table 1: PLCB1 Pathway Membership From the KEGG Pathway Database**

| Term | Ontology |
|---|---|
| intracellular signaling cascade | biological process |
| Learning | biological process |
| lipid catabolic process | biological process |
| Memory | biological process |
| oxygen and reactive oxygen species metabolic process | biological process |
| phosphoinositide metabolic process | biological process |
| signal transduction | biological process |
| calcium ion binding | molecular function |
| enzyme binding | molecular function |
| hydrolase activity | molecular function |
| phosphoinositide phospholipase C activity | molecular function |
| signal transducer activity | molecular function |
| intracellular signaling cascade | biological process |
| Learning | biological process |

**Table 2: PLCB1 Functional Annotations From the Gene Ontology.**

Failing to properly make the distinction between pathways and biological functions may result in misinterpretation of research results. One study concluded that outcomes of enrichment analyses differed based on the pathway database used for annotation due to differences in how pathway boundaries are defined between databases.[9] In essence, some databases tend to combine more molecular interactions into a single pathway than others, resulting in different results. Generalizing from this concept, boundaries around GO biological functions and

processes are likely to have little relation to pathway boundaries. Therefore, enrichment with biological function will *not* likely yield the same results as enrichment with pathways.

The following definitions will help bioinformaticians avoid confusion when communicating about enrichment analysis:

- *Pathway Enrichment Analysis*: Computational analysis of differential data sets using pathway annotations, typically from a pathway database such as KEGG or Pathway Commons.
- *Functional Enrichment Analysis*: Statistical analysis of differential data sets that using biological function or process annotations, typically from a database of biological functions, such as the Gene Ontology.

## CONCLUSION

Systems Biology is one of the most promising fields within translational research due to its unique ability to uncover the very mechanisms underlying human health. It is widely believed that a molecular pathway perspective will facilitate great strides in advancing medical treatment through the development of effective therapeutics targeted at specific molecules and interactions.

However, the complexity of this challenge is daunting. It requires combined effort from researchers in a wide variety of fields and data from many different sources. To ensure that research moves forward as rapidly and accurately as possible, all participating researchers must be familiar with the scientific domains associated with pathway research and must be consistent in their use of techniques and terminologies. Efforts to identify major research trends and to clarify evolving concepts are beneficial toward this goal.

Bioinformaticians play a key role in liberating meaningful information from vast quantities of raw experimental data by combining their knowledge of molecular biology with statistics and computer science. Bioinformaticians can develop tools and algorithms that accurately model and simulate pathway behaviors, efficiently enable access to databases of information, build the knowledge base for other researchers. In so doing, they can help make the compelling promise of systems biology and pathway research a reality.

# APPENDICES

## Appendix A: Literature Queries to Measure Growth of Pathway Research.

Articles in PubMed were searched using the following terms, limited to journal articles, and counted for each year 2000 through 2009, based on publication date:

| | | | |
|---|---|---|---|
| molecular pathway | molecular pathways | molecular network | molecular networks |
| metabolic pathway | metabolic pathways | metabolic network | metabolic networks |
| signaling pathway | signaling pathways | signaling network | signaling networks |
| signalling pathway | signalling pathways | signalling network | signalling networks |
| transduction pathway | transduction pathways | transduction network | transduction networks |

For example, the complete query string used to determine the number of relevant articles published in the year 2000 was:

("2000"[Publication Date] : "2000"[Publication Date]) AND ("signalling networks" OR "signaling networks" OR "transduction networks" OR "metabolic networks" OR "molecular networks" OR "signalling network" OR "signaling network" OR "transduction network" OR "metabolic network" OR "molecular network" OR "signalling pathways" OR "signaling pathways" OR "transduction pathways" OR "metabolic pathways" OR "molecular pathways" OR "signalling pathway" OR "signaling pathway" OR "transduction pathway" OR "metabolic pathway" OR "molecular pathway")

## Appendix B: Literature Queries to Measure Increasing Use of Bioinformatics Tools

Articles were further limited by the following terms commonly associated with bioinformatics activities, to estimate the significance of bioinformatics tools in supporting pathway research:

| | | | | |
|---|---|---|---|---|
| software | database | algorithm | bioinformatics | informatics |

For example, the complete query string used to determine the number of relevant articles containing both pathway analysis and bioinformatics terms published in the year 2000 was:

("2000"[Publication Date] : "200"[Publication Date]) AND (software OR database OR algorithm OR informatics OR bioinformatics) AND ("signalling networks" OR "signaling networks" OR "transduction networks" OR "metabolic networks" OR "molecular networks" OR "signalling network" OR "signaling network" OR "transduction network" OR "metabolic network" OR "molecular network" OR "signalling pathways" OR "signaling pathways" OR "transduction pathways" OR "metabolic pathways" OR "molecular pathways" OR "signalling pathway" OR "signaling pathway" OR "transduction pathway" OR "metabolic pathway" OR "molecular pathway")

## Appendix C: Articles Reviewed

The ISI Web of Knowledge was used to determine the three bioinformatics-related journals with the highest impact factors. Those journals were:

1. Briefings in Bioinformatics, Impact Factor 4.627
2. Bioinformatics, Impact Factor 4.328
3. BMC Bioinformatics, Impact Factor 3.781

Articles from these journals incorporating the term "pathway" in any field from the period April 15, 2009 through April 15, 2010 were reviewed. The table at the end of this appendix shows the results of this review. Columns are defined as:

- PMID: Unique PubMed identifier
- Article Title: Title of article as specified in PubMed
- Relevant: X denotes that the article was related to biological molecular systems in some way
- Focus (one or the other):
    - Method/Study: X denotes that the primary focus of the article was to document an experiment, study or method used in research.
    - Tool/Resource: X denotes that the primary focus of the article was to document a bioinformatics tool or resource; e.g., a downloadable computation algorithm or an online database.
    - Terminology (zero or more of the following):
        - "Pathway analysis": The literal string "pathway analysis" was used in any field in the article

        - "Network": The literal string "network" was used in any field in the article, to denote some type of molecular system. Examples of concepts *excluded* from these results would be Bayesian belief networks and computer networks.
    - Primary Activities as defined in Results and Discussion (one or more of the following):
        - Descriptive modeling
        - Predictive modeling
        - Curation
        - Pathway enrichment
        - Functional enrichment
        - Other

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMID | Article Title | Relevant | Method/Study | Tool/Resource | "Pathway Analysis" | "Network" | Descriptive Modeling | Predictive Modeling | Curation | Pathway Enrichment | Functional Enrichment | Other |
| 20385013 | web cellHTS: A web-application for the analysis of high-throughput screening data. | | | | | | | | X | | | |
| 20377890 | Gene set enrichment meta-learning analysis: next- generation sequencing versus microarrays. | X | X | | | | | | | X | | |
| 20374616 | Directionality in protein fold prediction. | | | | | | | | | | | |
| 20371497 | Payao : A Community Platform for SBML Pathway Model Curation. | X | | X | | | | | | | | |
| 20363732 | Streamlining the construction of large-scale dynamic models using generic kinetic equations. | X | X | | | | | X | X | | | |
| 20363729 | An Integer Programming Formulation to Identify the Sparse Network Architecture Governing Differentiation of Embryonic Stem Cells. | X | X | | | | | X | X | | | |
| 20359363 | GeneMesh: a web-based microarray analysis tool for relating differentially expressed genes to MeSH terms. | X | | X | | | | | | X | X | |
| 20356373 | New components of the Dictyostelium PKA pathway revealed by Bayesian analysis of expression data. | X | X | | X | X | | X | X | | | |
| 20353603 | Knowledge-guided gene ranking by coordinative component analysis. | X | X | | X | | | | | X | X | |
| 20335275 | PathWave: Discovering patterns of differentially regulated enzymes in metabolic pathways. | X | | X | X | X | | | | X | | |
| 20305269 | CoP: a database for characterizing co-expressed gene modules with biological information in plants. | X | | X | | | | | | | | |
| 20228128 | Pathway discovery in metabolic networks by subgraph extraction. | X | X | | | | | X | | | | |
| 20222969 | KiDoQ: using docking based energy scores to develop ligand based model for predicting antibacterials. | X | | X | | | | X | | | | |
| 20122237 | NeMo: Network Module identification in Cytoscape. | X | | X | | | | X | | | | |
| 20122228 | Virtual Screening of potential drug-like inhibitors against Lysine/DAP pathway of Mycobacterium tuberculosis. | | | | | | | | X | | | |
| 20122211 | Comparative classification of species and the study of pathway evolution based on the alignment of metabolic pathways. | X | X | | | | | | | | | X |
| 20122205 | Detection of characteristic sub pathway network for angiogenesis based on the comprehensive pathway network. | X | X | | | X | | X | | | | |
| 20122204 | Reaction graph kernels predict EC numbers of unknown enzymatic reactions in plant secondary metabolism. | X | X | | | X | | X | | | | |

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMID | Article Title | Relevant | Method/Study | Tool/Resource | "Pathway Analysis" | "Network" | Descriptive Modeling | Predictive Modeling | Curation | Pathway Enrichment | Functional Enrichment | Other |
| 20158919 | HDAPD: a web tool for searching the disease-associated protein structures. | X | | X | | | | | | | | |
| 20150321 | Fast and efficient searching of biological data resources–using EB-eye. | X | | X | | | X | | | | | |
| 20144236 | Simulation of a Petri net-based model of the terpenoid biosynthesis pathway. | X | X | | | X | | X | | | | |
| 20139469 | Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. | X | | X | | X | X | | | | | |
| 20109181 | Testing for mean and correlation changes in microarray experiments: an application for pathway analysis. | X | X | | X | X | | | | X | | |
| 20106820 | Predicting biodegradation products and pathways: a hybrid knowledge- and machine learning-based approach. | | | | | | | | | | | |
| 20070902 | Detecting disease associated modules and prioritizing active genes based on high throughput data. | X | X | | | X | | X | | | | |
| 20064243 | Modelling p-value distributions to improve theme-driven survival analysis of cancer transcriptome datasets. | X | X | | | | | | | | | X |
| 20064214 | Machine learning methods for metabolic pathway prediction. | X | X | | | X | | X | | | | |
| 20056730 | Comparative study of computational methods to detect the correlated reaction sets in biochemical networks. | X | X | | X | X | | X | | | | |
| 20055992 | SKPDB: a structural database of shikimate pathway enzymes. | X | | X | | | | | | | | X |
| 20043860 | Detection of gene pathways with predictive power for breast cancer prognosis. | X | X | | X | X | | | | X | | |
| 20031970 | Pandora, a pathway and network discovery approach based on common biological evidence. | X | X | | X | X | | X | | | | |
| 20021670 | Using mechanistic BaYesian networks to identify downstream targets of the sonic hedgehog pathway. | X | | X | | | | X | X | | | |
| 20021635 | A new permutation strategy of pathway-based approach for genome-wide association study. | X | X | | X | X | | | X | X | | |
| 20007742 | WebPARE: web-computing for inferring genetic or transcriptional interactions. | X | | X | X | | | X | X | | | |
| 20007251 | ChiBE: interactive visualization and manipulation of BioPAX pathway models. | X | | X | | X | X | | | | | |

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMID | Article Title | Relevant | Method/Study | Tool/Resource | "Pathway Analysis" | "Network" | Descriptive Modeling | Predictive Modeling | Curation | Pathway Enrichment | Functional Enrichment | Other |
| 19965882 | PathGen: a transitive gene pathway generator. | X | | X | | X | | X | | | | |
| 19965878 | The Gene Interaction Miner: a new tool for data mining contextual information for protein-protein interaction analysis. | X | | X | | X | | | | | | |
| 19955237 | Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. | X | | X | X | X | X | X | | | | |
| 19948066 | Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. | | | | | | | | | | | |
| 19933213 | Data integration for plant genomics–exemplars from the integration of Arabidopsis thaliana databases. | X | X | | | X | | | | | | |
| 19933158 | Pathway analysis using random forests with bivariate node-split for survival outcomes. | X | X | | X | | | | X | X | | |
| 19930714 | Identifying quantitative operation principles in metabolic pathways: a systematic method for searching feasible enzyme activity patterns leading to cellular adaptive responses. | X | X | | | X | | | | | | X |
| 19917102 | Algorithms for effective querying of compound graph-based pathway databases. | X | X | | X | X | | | | | | X |
| 19895694 | Elucidation of functional consequences of signalling pathway interactions. | X | X | | | X | | X | | | | |
| 19880371 | Simulation-based model selection for dynamical systems in systems and population biology. | X | X | | | X | | X | | | | |
| 19863822 | Inferring within-patient HIV-1 evolutionary dynamics under anti-HIV therapy using serial virus samples with vSPA. | | | | | | | | X | | | |
| 19828080 | Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering. | X | X | | | | | | | | | X |
| 19828078 | The Human EST Ontology Explorer: a tissue-oriented visualization system for ontologies distribution in human EST collections. | X | | X | | | | | | | | |
| 19828069 | Data recovery and integration from public databases uncovers transformation-specific transcriptional downregulation of cAMP-PKA pathway-encoding genes. | X | X | | | X | | X | | | | |
| 19840374 | TIde: a software for the systematic scanning of drug targets in kinetic network models. | X | | X | X | X | | X | X | | | |

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMID | Article Title | Relevant | Method/Study | Tool/Resource | "Pathway Analysis" | "Network" | Descriptive Modeling | Predictive Modeling | Curation | Pathway Enrichment | Functional Enrichment | Other |
| 19837719 | INTERSNP: genome-wide interaction analysis guided by a priori information. | X | | X | | | | | | X | X | |
| 19811690 | Computational analysis of gene expression space associated with metastatic cancer. | X | X | | X | | | | | X | | |
| 19811689 | HPD: an online integrated human pathway database enabling systems biology studies. | X | | X | | | | | | | | |
| 19811687 | Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. | X | X | | | | | | | | | X |
| 19811677 | Microarray platform consistency is revealed by biologically functional analysis of gene expression profiles. | X | X | | X | | | | | X | | |
| 19814801 | KEGGconverter: a tool for the in-silico modelling of metabolic networks of the KEGG Pathways database. | X | | X | | X | | X | | | | |
| 19808883 | ncRNAppi–a tool for identifying disease-related miRNA and siRNA targeting pathways. | X | | X | | | | | X | | | |
| 19808881 | Pathway identification by network pruning in the metabolic network of Escherichia coli. | X | X | | | X | | X | X | | | |
| 19796403 | Service-based analysis of biological pathways. | X | | X | | X | | X | | | | |
| 19793869 | Computing the shortest elementary flux modes in genome-scale metabolic networks. | X | X | | | X | | X | | | | |
| 19786482 | Integration of heterogeneous expression data sets extends the role of the retinol pathway in diabetes and insulin resistance. | X | X | | X | | | | | | X | |
| 19785734 | Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks. | X | X | | X | X | | X | | | | |
| 19761572 | Analysis of AML genes in dysregulated molecular networks. | X | X | | | X | | X | X | | | |
| 19758470 | From SNPs to pathways: integration of functional effect of sequence variations on models of cell signalling pathways. | X | X | | | X | | | | | | |
| 19758431 | Automated seeding of specialised wiki knowledgebases with BioKb. | X | | X | | X | | | | | | |
| 19744994 | Phenotypic categorization of genetic skin diseases reveals new relations between phenotypes, genes and pathways. | X | X | | | | | | X | X | | |
| 19736253 | Predicting homologous signaling pathways using machine learning. | X | X | | | X | | X | X | | | |

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PMID | Article Title | Relevant | Method/Study | Tool/Resource | "Pathway Analysis" | "Network" | Descriptive Modeling | Predictive Modeling | Curation | Pathway Enrichment | Functional Enrichment | Other |
| 19736252 | Efficiently finding genome-wide three-way gene interactions from transcript- and genotype-data. | X | X | | | X | | X | | | | |
| 19725948 | Comparative study of gene set enrichment methods. | X | X | | X | X | | | | X | | |
| 19703314 | GEOGLE: context mining tool for the correlation between gene expression and the phenotypic distinction. | X | | X | | | | | | | | |
| 19696047 | A boosting approach to structure learning of graphs with and without prior knowledge. | X | X | | | X | | X | | | | |
| 19696044 | Metabolite and reaction inference based on enzyme specificities. | X | | X | | X | | X | | | | |
| 19628508 | Path: a tool to facilitate pathway-based genetic association analysis. | X | | X | | | | | X | | | |
| 19628504 | PathBuilder–open source software for annotating and developing pathway resources. | X | | X | | X | X | | | | | |
| 19620097 | The SNP ratio test: pathway analysis of genome-wide association datasets. | X | | X | X | | | | | X | | |
| 19620095 | Caleydo: connecting pathways and gene expression. | X | | X | X | X | | | | X | | |
| 19608708 | An algorithm for learning maximum entropy probability models of disease risk that efficiently searches and sparingly encodes multilocus genomic interactions. | X | X | | X | | | X | | | | |
| 19578172 | Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model. | X | X | | | X | | X | | | | |
| 19570801 | Determining noisy attractors of delayed stochastic gene regulatory networks from multiple data sources. | X | X | | | X | | X | | | | |
| 19566964 | Clique-based data mining for related genes in a biomedical database. | X | X | | | | | | X | | | |
| 19563654 | NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. | X | X | | | | | X | | | | |
| 19563622 | Seeking unique and common biological themes in multiple gene lists or datasets: pathway pattern extraction pipeline for pathway-level comparative analysis. | X | X | | X | X | | | | X | | |
| 19561020 | Error control variability in pathway-based microarray analysis. | X | X | | X | | | | X | X | | |
| 19542154 | Reconstructing signaling pathways from RNAi data using probabilistic Boolean threshold networks. | X | X | | | X | | X | | | | |
| 19477995 | Network-based prediction of metabolic enzymes' subcellular localization. | X | X | | | X | | X | | | | |

| | | | Focus | | Terminology | | Primary Activities | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **PMID** | **Article Title** | **Relevant** | **Method/Study** | **Tool/Resource** | **"Pathway Analysis"** | **"Network"** | **Descriptive Modeling** | **Predictive Modeling** | **Curation** | **Pathway Enrichment** | **Functional Enrichment** | **Other** |
| 19473525 | GAGE: generally applicable gene set enrichment for pathway analysis. | X | X | | X | X | | | | X | | |
| 19426460 | TGF-beta signaling proteins and the Protein Ontology. | X | | X | X | | | | | | | |
| 19435746 | DIANA-mirPath: Integrating human and mouse microRNAs in pathways. | X | | X | | X | | | | X | | |
| 19432964 | Estimating parameters for generalized mass action models with connectivity information. | X | X | | | | | X | | | | |
| 19420052 | Approximate Bayesian feature selection on a large meta-dataset offers novel insights on factors that effect siRNA potency. | | | | | | | | | | | |
| 19414531 | PESTAS: a web server for EST analysis and sequence mining. | X | | X | | X | | | | | | |
| 19405941 | Identification of differentially expressed subnetworks based on multivariate ANOVA. | X | X | | | X | | | X | | X | |
| 19398450 | Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison. | X | | X | X | X | | X | | | | |
| 19336446 | Network-based multiple locus linkage analysis of expression traits. | X | X | | | X | | | | | X | |
| 19336445 | KiPar, a tool for systematic information retrieval regarding parameters for kinetic modelling of yeast metabolic pathways. | X | | X | | | | X | | | | |
| 19307239 | KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. | X | | X | | X | X | X | X | | | |
| 19289448 | Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data. | X | X | | | X | | X | X | | | |
| 19289442 | A concanavalin A-like lectin domain in the CHS1/LYST protein, shared by members of the BEACH family. | | | | | | | | | | | |
| 19286831 | Site of metabolism prediction for six biotransformations mediated by cytochromes P450. | X | X | | | | | X | X | | | |
| 19168911 | Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection. | X | X | | X | X | | | | X | | |
| **Count** | **100** | **92** | **56** | **36** | **27** | **53** | **6** | **43** | **21** | **20** | **6** | **7** |
| **Percent** | | **99%** | **56%** | **36%** | **27%** | **53%** | **6%** | **43%** | **21%** | **20%** | **6%** | **7%** |

## REFERENCES


1 Brater DC, Daly WJ. **Clinical pharmacology in the Middle Ages: principles that presage the 1st century.** *Clin Pharmacol Ther*. 2000 May;67(5):447-50.

2 Wainwright M, Swan HT. **C.G. Paine and the earliest surviving clinical records of penicillin therapy**. *Med Hist*. 1986 Jan;30(1):42-56.

3 **Systems Biology: the 21st Century Science**. *Institute for Systems Biology*. Available at http://www.systemsbiology.org/Intro_to_ISB_and_Systems_Biology/Systems_Biology_--_the_21st_Century_Science. Accessed May 12, 2010.

4 Ergün A, Lawrence CA, Kohanski MA, et al. **A network biology approach to prostate cancer.** *Mol Syst Biol*. 2007;3:82. Epub 2007 Feb 13.

5 Torkamani A, Topol EJ, Schork NJ. **Pathway analysis of seven common diseases assessed by genome-wide association**. *Genomics*. 2008 Nov;92(5):265-72.

6 King JY, Ferrara R, Tabibiazar R, et al. **Pathway analysis of coronary atherosclerosis.** *Physiol Genomics*. 2005 Sep 21;23(1):103-18.

7 **High Impact Journals**. Science Gateway. Available at http://www.sciencegateway.org/rank/index.html. Accessed May 12, 2010.

8 Ekins S, Nikolsky Y, Bugrim A, et al. **Pathway mapping tools for analysis of high content data**. *Methods Mol Biol*. 2007;356:319-50.

9 Green ML, Karp PD. **The outcomes of pathway database computations depend on pathway ontology**. *Nucleic Acids Res*. 2006 Aug 7;34(13):3687-97. Print 2006.

10 Arp R, Smith B. **Ontologies of cellular networks**. *Sci Signal*. 2008 Dec 16;1(50):mr2.

11 Faust K, Dupont P, Callut J, et al. **Pathway discovery in metabolic networks by subgraph extraction**. *Bioinformatics*. 2010 May 1;26(9):1211-8.

12 Huang Y, Li S. **Detection of characteristic sub pathway network for angiogenesis based on the comprehensive pathway network**. *BMC Bioinformatics*. 2010 Jan 18;11 Suppl 1:S32.

13 Salomonis, K Hanspers, AC Zambon, et al. **GenMAPP 2: new features and resources for pathway analysis**. *BMC Bioinformatics*, Jun 2007; 8: 217

14 **Signal Transduction Pathways**. Wikimedia Commons. Available at http://commons.wikimedia.org/wiki/File:Signal_transduction_pathways.png. Accessed May 12, 2010.

15 Gao B, Shen X, Kunos G, et al. **Constitutive activation of JAK-STAT3 signaling by BRCA1 in human prostate cancer cells**. *FEBS Lett*. 2001 Jan 19;488(3):179-84.

16 Sreenath SN, Cho KH, Wellstead P. **Modelling the dynamics of signaling pathways**. *Essays Biochem*. 2008;45:1-28. Review.

17 Vera J, Balsa-Canto E, Wellstead P, et al. **Power-law models of signal transduction pathways**. *Cell Signal*. 2007 Jul;19(7):1531-41.

18 Pitkänen E, Jouhten P, Rousu J. **Inferring branching pathways in genome-scale metabolic networks**. *BMC Syst Biol*. 2009 Oct 29;3:103.

19 Helleman J, Smid M, Jansen MP, et al. **Pathway analysis of gene lists associated with platinum-based chemotherapy resistance in ovarian cancer: the big picture**. *Gynecol Oncol*. 2010 May;117(2):170-6. Epub 2010 Feb 4.

20 Santos C, Eggle D, States DJ. **Wnt pathway curation using automated natural language processing: combining statistical methods with partial and full parse for knowledge extraction**. *Bioinformatics*. 2005 Apr 15;21(8):1653-8.

21 Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R. **Reconstructing signaling pathways from RNAi data using probabilistic Boolean threshold networks**. *Bioinformatics*. 2009 Sep 1;25(17):2229-35.

22 Shannon P, Markiel A, Ozier O, et al. **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res*. 2003 Nov;13(11):2498-504.

23 Babur O, Dogrusoz U, Demir E, et al. **ChiBE: interactive visualization and manipulation of BioPAX pathway models**. *Bioinformatics*. 2010 Feb 1;26(3):429-31.

24 Saraiya P, North C, Duca K. **Visualizing biological pathways: requirements analysis, systems evaluation and research agenda**. *Inf Visualization*. 2005; 1-15.

25 Spasic I, Simeonidis E, et al. **KiPar, a tool for systematic information retrieval regarding parameters for kinetic modelling of yeast metabolic pathways**. *Bioinformatics*. 2009 Jun 1;25(11):1404-11.

26 Waagmeester A, Pezik P, Coort S, et al. **Pathway enrichment based on text mining and its validation on carotenoid and vitamin A metabolism**. *OMICS*. 2009 Oct;13(5):367-79.

27 Matsuoka Y, Ghosh S, Kikuchi N, et al. **Payao: a community platform for SBML pathway model curation**. *Bioinformatics*. 2010 May 15;26(10):1381-3.

28 Kanehisa M, Goto S. **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res*. 2000 Jan 1;28(1):27-30.

29 Matthews L, Gopinath G, Gillespie M, et al. **Reactome knowledgebase of human biological pathways and processes**. *Nucleic Acids Res*. 2009 Jan;37(Database issue):D619-22.

30 Waagmeester AS, Kelder T, Evelo CT. **The role of bioinformatics in pathway curation**. Genes Nutr. 2008 Dec;3(3-4):139-42.

31 Abatangelo L, Maglietta R, Distaso A, et al. **Comparative study of gene set enrichment methods**. *BMC Bioinformatics*. 2009 Sep 2;10:275.

32 Subramanian A, Tamayo P, Mootha VK, et al. **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci USA*. 2005 Oct 25;102(43):15545-50.

33 Korde LA, Lusa L, McShane L, et al. **Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer**. *Breast Cancer Res Treat*. 2010 Feb;119(3):685-99.

34 Nie H, Neerincx PB, van der Poel J, et al. **Microarray data mining using Bioconductor packages**. *BMC Proc*. 2009 Jul 16;3 Suppl 4:S9.

35 **An Introduction to the Gene Ontology**. The Gene Ontology. Available at http://www.geneontology.org/GO.doc.shtml. Accessed May 12, 2010.

36 Soldatova LN, King RD. **Are the current ontologies in biology good ontologies?** *Nat Biotechnol*. 2005 Sep;23(9):1095-8.