EXECUTIVE SUMMARY
Understanding complex biological processes is critical whether doing basic research or trying to find the next blockbuster drug. Large quantities of information about these processes are available in the scientific literature, but finding the right information is both difficult and time consuming. By providing expertly-curated pathway diagrams based on validated data from the literature, Elsevier can give researchers working on a wide range of topics a jump-start to analyzing their experimental data, exposing some of the key processes of their particular system.
Introduction

Understanding the molecular processes that drive the initiation or progression of a disease is a difficult challenge. There is still a lot to learn about the complex interactions occurring between genes, proteins, and other molecules inside and outside of cells. Genomics technologies have increased the burden of sifting through vast quantities of data to find answers, but they have also allowed researchers to identify more connections between these various biological entities, which, in turn, leads to greater understanding.
The ability to map these connections helps researchers to understand the complex connected networks that can drive conditions like disease onset and progression, and can help them uncover the molecular components involved in the response to treatment. This knowledge can improve our ability to develop more useful diagnostics and more effective treatments.

But biological networks are complex, and assembling them into a more comprehensive picture of a disease is a task that requires specialized training and expertise. While the information that describes a biological pathway may be spread across hundreds if not thousands of scientific publications, few researchers have the time or resources to connect all the dots, and validate the resulting pathway maps.

Fortunately, Elsevier has a team of trained data curators who have set out to sift through the canonical knowledge and assemble a state-of-the-art knowledgebase from all that is currently known about molecular interactions, biological pathways and processes in humans, and to present that information in an accessible format for use by others in support of their biomedical research.

Elsevier’s Pathway Studio already contains more than 1,500 expertly-curated pathways, covering such diverse areas as signal transduction, metabolism, cell processes, neurobiology, expression targets, and drug toxicity. And for the past three years, these curators have been adding disease pathways to that collection. With a goal to cover as many diseases as possible, starting with the most well-studied, they have already built pathway collections for more than one hundred diseases, and they continue to add new pathways for at least five diseases per month.

Elsevier aims to provide reference pathways based on the most up-to-date information in the scientific literature for the hundreds of diseases known to humans. Ultimately, this vast, diverse collection could help biomedical researchers contribute to countless disease cures.
Andrey Kalinin and Sergey Sozin are among a team of Ph.D. scientists at Elsevier who are involved in the project. The resource they have established promises to contribute to furthering research into diagnosis and cure of diseases such as breast cancer, prostate cancer, various blood cancers, diabetes, Alzheimer’s, Parkinson’s, and other inherited, severe, and frequently occurring diseases.

Dr. Kalinin sums up the exhaustive process of building a disease pathway for inclusion in the collection: “In a nutshell, we find papers from experts in the field and reproduce the pathways based on those expert’s views, and then enter that analytical knowledge into Pathway Studio.”

But even the first step he describes is daunting: How to identify the experts in a given field of disease research? Kalinin says that status is determined by a quantitative analysis of most-referenced papers. He and his fellow curators give preference to “review papers” and those that describe experiments with human rather than model organism data. “Lots of papers talk about physiology of a disease, but we’re looking for papers that describe molecular mechanisms such as the role of specific genes/proteins and the mode of their interaction,” he says.

Figure 1. Basal cell breast cancer pathway.
Steps to building a pathway

Take for instance Mantle Cell Lymphoma, a rare type of non-Hodgkin lymphoma. Until very recently, Mantle Cell Lymphoma was a devastating diagnosis for any patient. But as a team of European researchers described last year in Nature, “unraveling different pathways of cell survival and progression” has led to “dramatically improved treatment activity” and “innovative targeted molecules” that have vastly improved the prognosis.

To build the pathway for Mantle Cell Lymphoma, Dr. Sozin says he searched PubMed, Elsevier sources, and Google (text and image searches) for key phrases such as “Mantle Cell Lymphoma Pathogenesis.” The search pulled up close to 30 full-text articles—some were open-access or available through Elsevier, while others came from other publishers. Elsevier has agreements with a wide range of scientific publishers so that their full-text content can be reviewed and mined alongside Elsevier’s content, ensuring a broad and unbiased review of the literature.

Sozin then mined those articles for crucial information about the classification and subdivision of the disease, the main pathways and de-regulated processes involved, graphical depictions of existing partial or complete disease pathways, and the most-referenced proteins for the disease. He also conducted key-word searches of widely used information resources like KEGG (www.genome.jp/kegg/) and, of course, Pathway Studio, for existing relevant pathways.

Kalinin says such a comprehensive search is typical of each pathway the team builds. “Our curators can read dozens of papers for a single disease,” he says. “It depends on how fast we find the good ones and how complete the information is across those papers.” Curators often scan additional fragments or excerpts of papers, book chapters, and “whatever else is available” to verify their findings, he says. “The main idea here is to find the papers that present the molecular mechanism of the disease in enough detail so that a coherent pathway can start to be built.”

Kalinin emphasizes that Elsevier’s researchers aim “to find out what the consensus in the scientific community is, not to develop theories ourselves.” He says, “We want to find recognized expert and make sure these experts express a formal opinion or one that is closest to the opinion that most people share.”

Sozin says a disease literature analysis lets him begin to draw the pathway, starting at a cell-changing signal, such as a mutation, translocation, ligand-receptor interaction, or over-or under-expressed gene. He transfers pictures and text from the top articles and draws a “sub-section” pathway that will be combined into the disease overview pathway.

For instance, he says, the sub-section pathways of Mantle Cell Lymphoma are:

- Deregulation of cell cycle and DNA repair
- Deregulation of apoptosis
- PI3K/AKT/MTOR, NF-κB and BCR signaling deregulation
- Canonical WNT signaling, and Hedgehog signaling

“Usually, I try to build a net with ligand receptors at the top, expression targets and cell processes at the bottom, and transmitter proteins in the middle,” Sozin explains. “If the signal is coming from a ligand receptor, it makes sense to build
the consequences of the molecular event from the top to the bottom. If the signal is coming from the mutated or fusion gene, it makes sense to build only downstream processes.”

From start to finish, the pathway-building process is manual. Kalinin says the only “semi-automated step” is to leverage Pathway Studio to find every protein ever associated with a given disease.

Sozin describes two outcomes of establishing relationships between those proteins. If the relations already exist in Pathway Studio, he chooses the strongest and most relevant ones, checks for false positives, and confirms that the effect—positive or negative—is correctly indicated there. On the other hand, if he finds no relation in Pathway Studio but does find one in PubMed, he creates a new relation and links to Medline, transferring all of the information for every protein from the articles to Pathway Studio.

When gaps remain in his net of proteins, Sozin says he searches the Internet and Pathway Studio for neighbors of each protein. “I try to confirm the role of every protein on the pathway in the disease pathogenesis.”

Sozin says he uses common and well-known transmitters to fill in gaps in the signal transduction chain. To complete the Mantle Cell Lymphoma pathway, for instance, he searched “Mantle Cell Lymphoma + protein name” for every protein on the pathway. The goals, he says, are to reproduce the consequence of pathological events as a “net” structure and to put as much “canonical” confirmed knowledge into the pathway as possible.

Once sub-section pathways for a given disease are built, curators build an overview pathway from those smaller parts. “Sub-section pathways are much easier to analyze, but the overview pathway gives a more general and comprehensive perspective,” Sozin says. Curators also write short narratives for every pathway, which might include information from the literature such as drugs that exist to target the pathway. This information helps users understand the context for the newly mapped pathways.
Contributing to the molecular biology canon

Kalinin notes that the value of collaborations among the curators on his team, each of whom has become an expert in the disease domains they cover, cannot be overstated. “Many diseases use similar cascades. When someone builds new disease pathways that are from an area similar to one that another curator has already covered, they can get instant verification from a colleague as to whether the pathway makes sense.” This process broadens the expertise of the reviewers as well.

In turn, their work will enable researchers worldwide—even those who are not experts in a particular disease—to easily access and benefit from the current body of knowledge. “Even users with zero experience in a given disease area can get a general idea about a disease process from Pathway Studio,” Kalinin says. “They can see the proteins that are the major players, insert existing drugs to see the mechanism of their action, and they can identify and substantiate new targets for further drug development.”

Pathway Studio users can leverage the disease pathway collection as a reference about genes, diseases, or processes. They can map their experimental data to the pathways, to better understand their results in the context of previously published work. They can also add, subtract, and combine pathways or otherwise edit them for their own needs. If they disagree with Elsevier’s reference pathway, they can edit it and save a modified version on their computer. And they can combine them into bigger, more complex disease models.

Kalinin points to the value of access to the comprehensive pathway collection available in Pathway Studio. “Say you discover a new protein or a new gene and you want to find out what role it plays. You make a knockout mouse and study the expression of genes in that versus regular mice. By analyzing the differences in gene expression against our pathway collection, you can instantly find pathways relevant to the role of the gene in question.”

While Sozin notes that “any pathway is an approximate scheme reflecting the current knowledge,” Kalinin says the end result of the team’s hard work is a conservatively curated and validated list of every protein connected to a given disease—“a complete picture of what is described in all of the scientific literature since the beginning of time.” Exploring these reference pathways presents an efficient way for a researcher to get a sound foundation to understand the biology of a disease, and to build on that foundation with their own work, potentially extending the knowledge for everyone.

We are always interested in hearing from domain experts with comments or suggestions about how to improve our pathways. Please contact us at: info pathway@elsevier.com
Visit elsevier.com/products/solutions/pathway-studio or contact your nearest Elsevier office.

ASIA AND AUSTRALIA
Tel: +65 6349 0222
Email: sginfo@elsevier.com

EUROPE, MIDDLE EAST AND AFRICA
Tel: +31 20 485 3767
Email: nlinfo@elsevier.com

JAPAN
Tel: +81 3 5561 5034
Email: jpinfo@elsevier.com

NORTH AMERICA, CENTRAL AMERICA AND CANADA
Tel: +1 888 615 4500
Email: usinfo@elsevier.com

KOREA AND TAIWAN
Tel: +82 2 6714 3000
Email: krinfo.corp@elsevier.com

SOUTH AMERICA
Tel: +55 21 3970 9300
Email: brinfo@elsevier.com