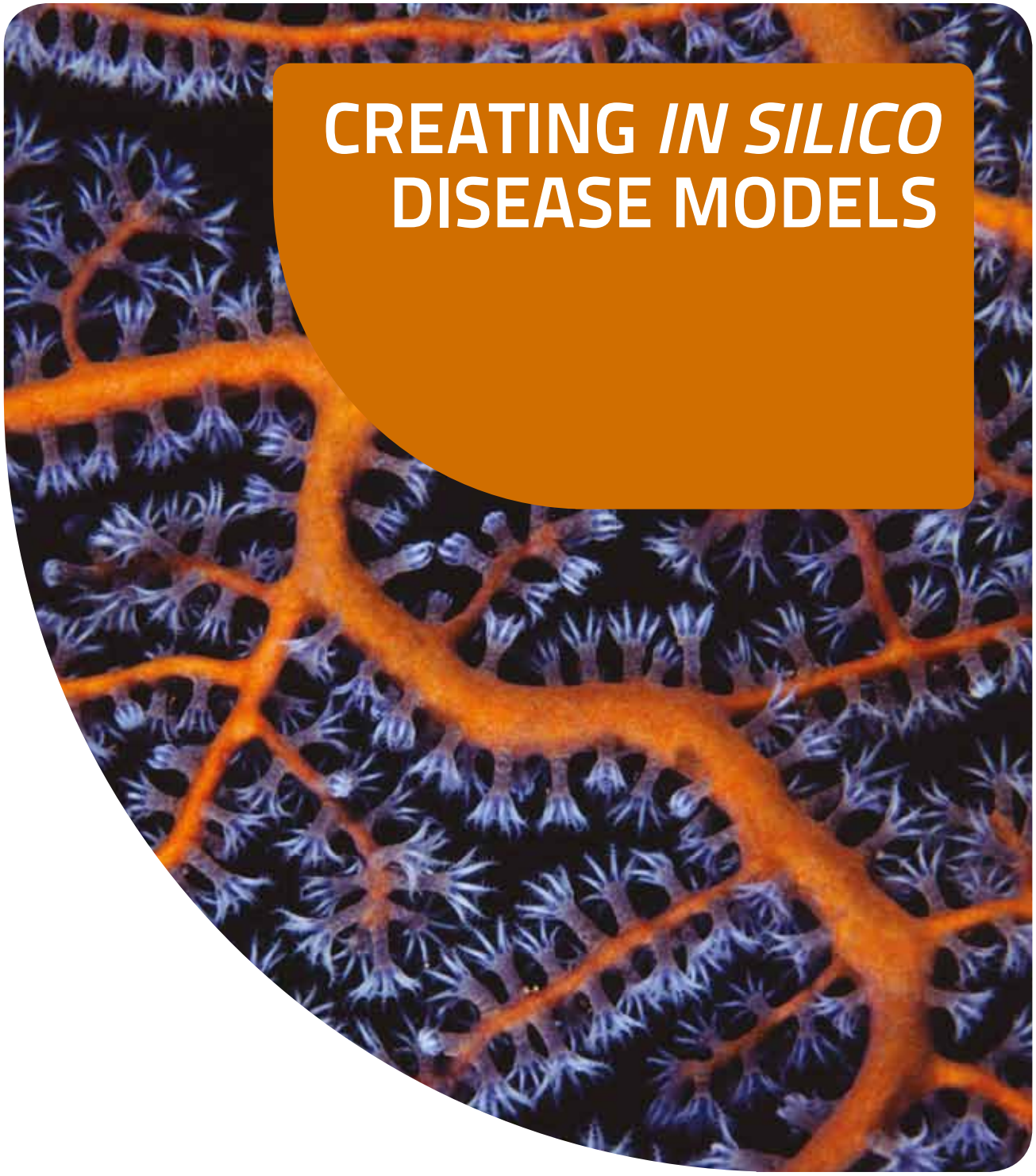
ELSEVIER



CREATING *IN SILICO* DISEASE MODELS

PATHWAY STUDIO®

THE "CONNECTIONIST" APPROACH TO FINDING DRUG TARGETS

The pharmaceutical industry has never been under greater pressure to bring new, effective and profitable drugs to market. Yet the R&D process is broken: The cost to bring new therapies to market soars while the number later withdrawn due to safety concerns mounts. Meanwhile, genomics provides new insights into diseases, but also reveals more biological complexities than had ever been imagined. Discoveries that multiple factors contribute to disease progression and treatment response, or that interactions among diseases affect the course of each one, have dramatically intensified investigatory workloads.

In this environment, identifying and validating drug targets requires evidence from many sources. Studies that illuminate the functions of individual proteins and examine their interactions with other proteins and processes are indispensable to these complex investigations.

The scientific literature is full of thousands upon thousands of such studies. They describe discoveries about proteins involved in the onset, progression and pathogenesis of diseases that have been made by experimenting on protein modulation, investigating cellular-level assays, profiling pathways, creating animal disease models and analyzing genome-wide associations.

Having the ability to link proteins of interest to other proteins, similar diseases, relevant biological functions or existing drugs, lets a researcher draw a more complete picture of a disease or process. These "connectionist" models are highly useful in formal statistical reasoning. Even simply identifying all of the proteins connected to a disease provides a starting point for a more thorough investigation.

Knowing how many times a relationship has been cited in the scientific literature between each protein and a given disease, for instance, could be the basis for prioritizing proteins for a closer look. A project might focus on those with well-established links or pursue riskier, but potentially novel targets: those with weaker or more recently discovered connections.

MINING THE SCIENTIFIC LITERATURE

Knowing this information is in the literature is one thing - finding it is another. Traditionally researchers have relied on reading abstracts, or conducting keyword searches of PubMed to find literature on their topics of interest. However authors are notoriously varied in the way they describe their research, so it's easy to miss some of the possible search terms, which can lead to missing critical publications.

Instead of manually combing through hundreds of scientific papers to find precisely relevant research, automated literature mining systems can do a more rapid and more thorough search of the literature, ensuring that little evidence is missed. Users select specific genes, proteins, drugs or types of relationships of interest and these systems point them to the handful of papers that deserve deeper attention. Researchers are thereby able to quickly identify areas of interest, avoid information overload and more rapidly develop and confirm hypotheses.

Elsevier has a unique approach to helping researchers retrieve relevant data from the scientific literature in search of potential new drug targets. Our solution for biology, Pathway Studio, focuses on pathway analysis and is powered by proprietary natural language processing-based (NLP) text-mining technology. Using NLP, we have extracted and made readily available crucial information about genes, proteins, small molecules, metabolic and disease processes, and a variety of other interaction data from more than 3 million full-text articles and 23 million PubMed abstracts.

Our tool extracts sentences that contain key concepts, such as "gene," "protein" or "drug," and then reviews the sentences for relationships described by phrases like "binds to," "activates" or "inhibits." Those sentences, as well as biological pathways derived from that information and validated by Elsevier experts, are loaded into the Pathway Studio knowledgebase for researchers to search, view and apply to analyze their experimental data in the context of previously published observations.

Pathway Studio enables investigators to filter out reams of data that are not related to the disease or phenotype of interest, as well as to zero in on the proteins, genes, diseases, drugs, and other entities that the data-mining process identifies as relevant.

Scientists can also use this solution to build and visualize complex "in silico" models to represent the biological processes of a disease or a drug response. They can use such models both to direct and to evaluate the outcome of their laboratory studies.

NETWORK-BASED DISEASE ANALYSIS

One approach is to identify connections between a disease of interest and other closely related diseases and biological processes. For instance, if genome-wide association studies (GWAS) have shown a given protein to be involved in a disease, it might be useful to look at that protein's role in relevant biological processes. Greater confidence can be assigned to those proteins that have more associations to various disease-associated processes.

In one example of this kind of network-based disease analysis, one of our customers was able to identify targets for drugs to treat acute heart failure (AHF). Their starting point was an existing drug targeting AHF. They verified a drug target, the disease mechanisms and the pathways it impacts, the immediate network neighbors, and associated proteins and biological processes. This let them build a credible disease model.

As a starting point, the researchers searched our biology database for the mechanism of action of another drug currently in clinical trials, which targets the relaxin receptor. The search revealed 56 diseases related to AHF including hypertension, coronary artery disease, cardiomyopathy, ischemia, atrial fibrillation and infarction. It also turned up 36 closely associated biological processes including blood pressure, blood circulation, vasodilation, endothelial cell function, vasoconstriction and coronary flow.

To find pathways that might be targeted to treat AHF, the researchers looked at all of the proteins associated with those diseases and biological processes. Several biologically-relevant potential targets were identified, and then ranked for further investigation. Ultimately, four pathways—two novel, as well as two suspected—were identified as possible alternative AHF targets, and were the focus of in-house validation studies.

PROTEIN-PROTEIN INTERACTIONS

Another approach to associating proteins with diseases and ranking their significance or validity as drug targets is to compare their expression in diseased versus normal tissue. When done using Elsevier's biology solution, this practice might yield hundreds or even thousands of differentially expressed genes that can be quickly ranked according to the strength of their relationship to the disease. The database search might directly or indirectly link them to the disease of interest, or it might show them to have strong associations with each other. This provides a starting point for further analysis, saving the researcher many hours of manually reviewing the literature.

The ability to tie multiple differentially expressed proteins to each other, as well as to a disease or disease-associated process, gives researchers greater confidence in their role in the disease mechanism. For instance, groups or clusters of tightly interconnected proteins could point to connections between pathways relevant to a disease. To estimate the specificity of proteins to a particular pathway, researchers can compare the number of connections among proteins in that pathway to the total number of relationships of any type for those proteins. This connectionist approach offers semantic richness and may suggest multiple selection criteria.

DEVELOPING DISEASE BIOMARKERS

Investigators can also employ the connectionist approach to stratify patients in support of personalized treatments. By augmenting formal network models with protein ontologies, investigators can identify and define the specificity of disease biomarkers. They can use Elsevier's biology solution to identify proteins directly downstream from transcriptionally modulated processes by focusing on transcription factors associated with that disease, and then on downstream secreted molecules, which could be examined as potentially measurable biomarkers.

Profiling experiments that compare normal tissue to diseased tissue—yielding long lists of differentially expressed genes—can also be used to identify both candidate drug targets and biomarkers. With a formal data model, investigators can quickly rank proteins by connecting them to each other and to diseases or disease-related biological processes.

Going further, investigators can apply statistical causal reasoning algorithms to the protein network to identify key transcriptional regulators that become activated in disease and modulate proteins. These regulators might otherwise be overlooked in a traditional expression-based ranking because they do not themselves change expression.

In one example of a use of Pathway Studio to study biomarkers, researchers at Elsevier identified a potential biomarker for drug-induced cholestasis. The researchers started by building a cholestasis network model based on unconnected data in the literature. First, they searched the biological networks database to compile a list of drugs that are known to induce cholestasis and a list of proteins shown to be either directly involved in the development of cholestasis, or frequently reported to be mutated in cholestasis patients. Then, from the drug-protein relationship database, they pulled a list of proteins modulated by cholestasis-inducing drugs.

From those lists, the researchers created a single list of all possible protein interconnections. They ranked those according to proportion of connections to each other versus total connections. Next, they ranked them by the number of times the literature supported a connection to cholestasis and by the number of drug-protein connections.

The resulting cholestasis model immediately revealed two key transcription factors implicated in cholestasis—FXR and PXR, which are activated by bile acid accumulating in hepatic cells. Further searches in the network database identified secreted proteins that are regulated by both factors, leading to the discovery of FGF19 as a possible biomarker for drug-induced cholestasis. Since publication, multiple independent studies have confirmed results predicted by the model. Most notably, a subsequent experiment confirmed that FGF19 levels increase during cholestasis.

The scientific literature already contains vast amounts of information about genes, proteins, diseases and biochemical processes. The challenge is sifting through all those publications to find the specific facts relevant to an individual research project. Elsevier's biology solution helps researchers mine and apply that information to build models to expand our understanding of disease processes. Examining networks, ranking protein-protein interactions and conducting profiling experiments to identify biomarkers are a few ways investigators can use this solution to build in silico predictive disease models that are invaluable in guiding laboratory experiments and improving the chances of identifying novel drug targets.

In an increasingly costly, competitive and complex pharmaceutical R&D environment, our tool empowers investigators with sophisticated approaches to understanding disease mechanisms, detecting drug targets and identifying biomarkers. Researchers enabled by Pathway Studio can more quickly identify areas of interest, create in silico models to more rapidly develop and confirm hypotheses, and reach their goals with greater confidence.

Learn more about Elsevier's biology solution here.

ELSEVIER

FURTHER INFORMATION

Please vist www.elsevier.com/pathway-studio

TECHNICAL SUPPORT/E-CUSTOMER SERVICE

AMERICAS
(8am-8pm CET - St. Louis)
Email: usinfo@elsevier.com
Brazil: brinfo@elsevier.com
US toll-free: +1 (888) 615 4500

JAPAN
(9,30am-5,30pm JST, Tokyo office)
Email: jpinfo@elsevier.com
Website: japan.elsevier.com
Tel: +81 (3) 5561 5035

ASIA AND PACIFIC
(9am-6pm SST, Singapore office)
Email: sginfo@elsevier.com
Tel: +65 6349 0222

EUROPE, MIDDLE EAST AND AFRICA
(9am-6pm GMT+1, Amsterdam office)
Email: nlinfo@elsevier.com
Tel: +31(0) 20 485 3767

PATHWAY STUDIO®