Customer case study
Field of study: Core Lab

The Big Picture:
From Data to Biological Knowledge

Dr. Michael Edwards is always looking for the big picture. A biologist by training, he gravitated toward computational biology and bioinformatics as he encountered technologies that generated more data than anyone knew how to handle.

Now an assistant professor at the University of Colorado, Denver, Edwards was in graduate school when microarrays first came out. During his PhD studies at the University of Wisconsin, he remembers a paper published by the lab — one of the first major articles using this new tool for the study of longevity effects of caloric restriction — that included no statistics beyond fold changes. With gene chips, he saw immediately, "you're looking at a lot of information, and how do you make sense of it?"

Edwards completed a postdoc in the proteomics field, another realm of biology that was seeing the beginning of massive quantities of data, before coming back to gene expression and joining the University of Colorado's pulmonary division in 2007. His expertise in handling enormous data sets has made him the go-to collaborator for teams that generate long gene lists with no clear path forward. When this happens, Edwards has his own go-to expert: Ingenuity Pathway Analysis (IPA) from QIAGEN. "We're able to measure a lot of things, so the challenge is figuring out how to use all of this information to get to the big picture," he says. "That's what IPA allows me to do: bring the biology into the data."

First Encounter

Before IPA was released, Edwards found himself trying to make sense of these data sets with the most rudimentary tools. "When I first started using these gene arrays, we didn't even know what the probes were measuring. We had to BLAST search against the genome to find out what genes we were even looking at," he says. Once he did find out what genes were involved, he had to figure out what they did and how they were related to other genes differentially expressed in the same experiment. "I did this by hand, combing through PubMed for hours or days to look for relationships," he says.

As a graduate student, Edwards participated in a project using microarrays to examine gene expression changes taking place in skeletal muscle as a mammal ages, focusing on how caloric restriction affected the process. "As I was manually going through the data, I found all these genes related to p53," he recalls. The finding was unexpected, so Edwards took the time to track down the related genes, grouping them by genes that bind to p53, genes that p53 is known to induce, genes that p53 was known to inhibit, and so on. "They all reflected a pattern that suggested a more active p53 than we expected," he says. The process took weeks, and when it was completed, his PI didn't immediately buy into the importance of the groupings. "It took a lot of convincing to get him to believe that this was real," Edwards says.

When IPA was released, Edwards realized immediately that the tool would accomplish this type of endeavor for him — no combing through databases, no spending hours reading papers and trying to connect the dots by hand. "Ingenuity came along and basically had a database that would do this for me," he says. The tool's upstream regulator function was especially useful for extracting value from gene expression data. "IPA really simplifies this process and saves me a lot of time," he

adds. "It's not just a search of PubMed; you're searching the Ingenuity Knowledge Base, which is immense and has lots of relationships that you might never find with just a regular literature search."

The Peripatetic Professor

With Edwards' focus on interpreting biological data, his collaborations with experimental teams allow him to investigate several diseases at any one time. That suits his personality — and provides for the occasional serendipitous discovery.

In a recent project, Edwards worked with scientists who had gathered a great deal of information from sequencing bladder tumors. They had emerged with a list of some 425 mutated genes, representing the most extensive list of mutations for this type of cancer, but it was unclear how to proceed. How did all these genes fit together? What was the common theme? The sheer number of genes made the idea of interpretation quite daunting.

"They came to me with their gene list, and I put it into IPA, which started building pathways and likely signaling avenues," Edwards says. "The genes started to group based on biological functions — chromosome structure or cell cycle maintenance, for example — and they would build these networks within themselves." He and his collaborators went back to look at the tumors and found that key components of cell signaling would have to be mutated in order to produce a bladder tumor. "They were amazed," he says. "I get that response quite a bit. People don't really know that there's software out there that can do this." Without IPA, he adds, this project would have ended with publication of the list of 425 genes and no information about relationships between genes or how they function to create tumors.

As co-director of the Biostatistics, Informatics, and Bioinformatics Core in the UC Denver Lung SPORE (NIH's designated Specialized Program of Research Excellence), Edwards works closely with lung cancer researchers. He spends quite a bit of time helping scientists who have a gene or list of genes and need help turning that into something actionable. To avoid the pitfalls of false positives, he relies on two large lung cancer data sets and correlates expression data for the scientist's target gene or genes of interest across them. "I use two independent data sets to find what's commonly correlated, and use that to bring out some of these relationships: how do these related genes connect with the target gene, what pathways do they interact with?" The double database approach helps to filter out false positives and highlight genes that really are behaving like the target gene. "Once we have a collection of those genes, I can use IPA to interrogate that and find out what pathways might be overrepresented," Edwards says. "Is there a particular master regulator that I could modulate to affect the target gene activity?"

Investigating several different diseases works to his advantage, Edwards says. For example, as a thought exercise he used IPA to overlay results from a lung cancer study with results from a bladder cancer project. "It was

> "That's what IPA allows me to do: bring the biology into the data."

almost a perfect fit, suggesting commonality between these two different types of cancers," he notes. That moment of serendipity will require follow-up studies, and would not have happened without a software tool that made comparing the results so simple.

Inside IPA

As an IPA power user, there are several features that Edwards finds particularly useful for his research. Most important, he says, is that the tool lets users step away from the concept of an "idealized" pathway and toward something that has more biological meaning.

"What we've found out as you analyze gene expression is that known pathways have been built from information derived from many different experiments in many different cell types and organisms; they're idealized pathways, and depending on what you're looking at they might not necessarily apply," Edwards says. "In IPA you can group genes based on whether they had any interaction in the scientific literature, making your own pathways. IPA allows you to find the true cellular signaling networks without having to use an idealized pathway that might not even be applicable for what you're doing."

Other features that make his life easier include Causal Network Analysis and multi-hop, which help reveal upstream and indirect regulators. "What a lot of us in bioinformatics are finding is that some of the truly important events in cellular signaling are probably invisible in the transcriptome," he says. Tracking down these master regulators based on transcriptome data is possible with IPA and was "a huge step" in accelerating Edwards' work. "These tools are very good at getting at the skeleton that connects all of the gene expression data," he adds. Edwards has found IPA very easy to use — in fact, so easy that he is now using it to teach bioinformatics to high school students. He is just kicking off a program with a school in an underserved Denver area in which students will import array data from free public databases and then analyze it using tools including IPA. His goal is to help kids connect with biology and make a positive impression.

"Biology is quickly turning into numbers. What IPA allows you to do is to pull all that information together and give you a big picture of what's going on," Edwards says. "You can also zoom in and find small things. With IPA, I find these relationships that aren't even in the scientific literature and that I never would have guessed occur."

QIAGEN Silicon Valley
1700 Seaport Blvd. Third Floor
Redwood City, CA 94063
Tel. +1 650 381 5100
Fax. +1 650 381 5190
info@ingenuity.com
www.ingenuity.com