Ansuman Chattopadhyay, PhD, directs the Molecular Biology Information Service offered by University of Pittsburgh's Health Sciences Library System.
Photo by Rob Larson

Researchers harness the power of the Illumina Correlation Engine

Investigators lean on the comprehensive database to accelerate discovery and publication

RESEARCHERS OFTEN NEED greater scientific context to help them complete their studies. How do their results fit in with other people's work? What does the literature say about a gene's expression? Are there available drugs that modulate a specific gene?

For Ansuman Chattopadhyay, PhD, who directs the Molecular Biology Information Service offered by University of Pittsburgh's Health Sciences Library System,*[1] helping researchers find that biological context is essential. By guiding investigators to the most appropriate resources, he helps them gain new insights into how their work fits into global efforts to understand biology and disease, and how to move their research forward. In turn, this data provides ammunition to procure more grants and support publication in more prestigious journals.

Public databases hold an enormous amount of contextual data, but they can be difficult to use. In some cases, searching for context can become its own time-consuming project. The National Center for Biotechnology Information's Gene Expression Omnibus (GEO),[2] a public database for array and sequencing data, is a good example.

"GEO is a gold mine," says Chattopadhyay. "We can search the database to find how genes are expressed under certain conditions—different diseases or drug treatments or certain mutations. It's a wonderful resource, but the raw data is not immediately actionable. Researchers must make those comparisons themselves."

That's where the Illumina Correlation Engine (formerly BaseSpace Correlation Engine) comes in. This omics research database and suite of tools including Body Atlas, Disease Atlas, Pharmaco Atlas, Knockdown Atlas rapidly identifies public information in GEO and other databases. The information has been curated and normalized, giving the data additional power to help researchers define associations by tissue, disease, compound, and/or genetic perturbation.

For Chattopadhyay and colleagues, Correlation Engine has been a windfall, helping them understand gene function, drug activity, and other mechanisms, and contextualizing their results for publication and grant applications.

Moving research forward
Correlation Engine has been an invaluable tool for

For Research Use Only. Not for use in diagnostic procedures.

researchers at universities, providing essential data and insights to get projects over the finish line. The software has been particularly useful for understanding how genes are over- or under-expressed in certain conditions.

"Thousands of genes may be upregulated and thousands may be downregulated, but what does that mean?" Chattopadhyay asks. "We can upload that dataset to Correlation Engine, which can identify datasets that positively or negatively correlate with the experimental work, and that's new information and context."

Correlation Engine can show scientists whether they are moving in the right direction or not, and source additional data points for validation, helping them make the case for publication.

These capabilities can also help identify drugs that modulate disease. During the early days of the pandemic, before vaccines were available, many scientists were investigating whether existing drugs could be repurposed to fight SARS-CoV-2. One study, with a big assist from Correlation Engine, identified 56 drugs with potential therapeutic action against the virus.[3]

"We found a good candidate, but after the vaccine came out, interest was low," says Chattopadhyay. "However, the study nicely captures the power of Correlation Engine, and we use that paper to train our researchers on the tool. It really highlights the tool's ease of use."

Confirming better models
Discovery research is often conducted in animal models; however, rodent biology and primate biology are quite distinct from human biology. To overcome this, scientists have been researching organoids, or human tissue grown *in vitro* that replicates the basic function or structure of an organ on a small scale. As a more accurate model of human biology, organoids could dramatically advance

discovery research, and Correlation Engine can help scientists understand how close these models are to *in vivo* human tissue.

"If you did a transcriptome study on an organoid, and you upload that and find data that's positively correlated with the actual tissue, that's good evidence the organoid is close to the tissue being modeled," says Chattopadhyay. "This has been particularly helpful in brain research."

The power of training
Chattopadhyay's team is impressed with Illumina's data curation and ongoing support. He says, "If we find data that is not present in Correlation Engine, we just request it and, within a couple of days, they make it available." Not every study passes the rigorous curation process, of course, but every week eight to 20 new studies are added, some via customer request.

The product has been a fixture of the library's licensed software collection for more than 10 years—even before Illumina acquired the technology—and around 700 people have registered to use it. Chattopadhyay's team offers regular classes, applying Correlation Engine to data described in published papers—like the one that identified existing drugs as candidates to treat SARS-CoV-2 infection—to demonstrate the tool's capabilities. For some students, the classes are their first experience using Correlation Engine. Others attend the workshops to advance their understanding of how they can use it in their research.

"I think the secret sauce for our success is that we offer the workshops," says Chattopadhyay. "Unfortunately, some universities just license a product, but it hardly ever gets used. I feel the training helps people get a better handle on the tool's capabilities. They realize they can apply this technique to their own research and make enormous progress." ◆

3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7336709