

Army of Five: How a Lean Core Lab Performs Mega-Scale Data Analysis and Biological Interpretation

At the University College London genomics core, scientists trust their data analysis to Ingenuity Variant Analysis and IPA and get to their next experiment even faster.



In the rapid-fire world of genomics, it's easy for researchers to feel that they are only as good as their next hypothesis. For scientists who work with the genomics core facility at the University College London, that's just fine — after all, they have a clearer path to their next hypothesis than most.

The UCL Genomics lab provides analytics for many types of experiments through Ingenuity® Pathway Analysis (IPA®) and Ingenuity® Variant Analysis™, web-based applications that allow users to rapidly analyze and interpret the biological significance of their genomics data. Mike Hubank, scientific director of UCL Genomics and senior lecturer at the Institute of Child Health, says that these tools utilize Ingenuity's massive Knowledge Base of biological findings to create mechanistic hypotheses that can help direct scientists to their next relevant experiment. "The real benefit is the ability to connect things you wouldn't otherwise have thought about," Hubank says. "It's a good tool for hypothesis generation, and there's a lot to be said for that."

Hubank knows a thing or two about the power of analytics. As head of a very active genomics core, he leads a staff of five — and each day has to find new ways to maximize what they can accomplish. After all, in addition to providing core lab services for a major university, the team also works with faculty at the Great Ormond Street Children's Hospital, a leading pediatric hospital in London. "We have hundreds of projects going on during a year, from basic research in many different areas to applied research, particularly in clinical and hospital settings," Hubank says. "We have to oversee a lot of data to do so." In the past, his team used free databases and open-source tools for their analysis needs, but they've learned that there's always an underlying cost to these options.

The genomics core's data primarily comes from next-generation sequencing instruments as well as gene expression and genotyping arrays. The data they

generate serves as the foundation for genome-wide association studies, exome characterizations, and targeted sequencing projects that are focused on human health. "One of the things we really struggle with is analytical capacity, which is where Ingenuity has come in so handy," Hubank says.

TIME SINK

In Hubank's core lab, data analysis can quickly become the biggest bottleneck. The team churns through nearly 3,000 Illumina genotyping arrays per month, and is on track to complete about 500 exome sequence analyses this year — not to mention all the other work going on. While the team of five can crank out the genomic data at industrial scale, analyzing all that information is a different story. "We don't have enough staff here to answer all the questions and do all the interpretations for everybody," Hubank says.

To address this challenge, he turned to Ingenuity Systems. Hubank, a longtime user of Ingenuity IPA, tried out the newer Variant Analysis application in hopes that it would help his team perform faster analyses of all the data they were processing for users. "I was quite surprised when we trialed it," he says, noting that it helped him to assess candidate lists of variants and identify those likely to be causative much faster than any other approach he has tried.

"Traditionally, once you've found a list of candidate variants, then you've got to sit there and look at the literature and try to figure out which you think are the most likely ones. That takes a long, long time even for an individual project," Hubank says. "Variant Analysis cuts out a lot of the time. It was a no-brainer to adopt the tool," he adds.

Today, the UCL Genomics team can generate the experimental data, boil that down to the differences between conditions being tested, and share the data with customers using Variant Analysis for

further interrogation. "Variant Analysis is a very nice solution for us because we can give users access to this very handy tool which we can easily teach them to use, and it will help them interpret and understand their data better," Hubank says. With these Ingenuity tools, Hubank's small staff has the analytical power of a virtual army of bioinformaticians.

"Variant Analysis cuts out a lot of the time. It was a no-brainer to adopt the tool."

Mike Hubank, scientific director of UCL Genomics and senior lecturer at the Institute of Child Health.

Variant Analysis has also helped Hubank in another dimension of running his core lab: figuring out what to charge for analysis. As a completely self-sufficient core lab, UCL Genomics has to cover 100 percent of its costs — but the question of how to charge for analysis using open-source tools was always a challenge. "Charging for the analysis part is tricky because we simply don't know how long a particular project is going to take," Hubank says. Now, he can charge a flat rate for access to Variant Analysis, and it's a much cleaner approach on the administrative side. "Using Ingenuity takes a lot of pain out of that chargeback process for us," he says.

THE COST OF FREE

Like so many core labs, UCL Genomics historically relied on open-source algorithms, pipelines, and databases to fill its analysis needs. With so many of these free tools available, why would Hubank voluntarily choose a paid solution instead?

"What you think you're getting free isn't actually free," says Hubank, whose team still uses open-source tools such as R, SIFT, and PolyPhen. So-called free tools still require people to develop the pipelines, maintain and update them, keep databases current, and more. "Someone has got to do that, and that someone has to be paid for," he says.

Indeed, the cost of free tools goes beyond the individuals in each lab tasked with their proper maintenance. Hubank remembers using well-known, highly regarded public databases for various analyses — only to find out after the fact that the database hadn't been kept current, and that the analysis would have to be thrown out because it didn't take critical new information into account. "Often you're using an out-of-date product without even knowing that it's out of date," Hubank says.

Thanks to many years of experience with IPA and his more recent use of Variant Analysis, Hubank has come to the conclusion that "you get what you pay for with analysis." He says that even among core lab clients who initially balk at having to pay for analysis, "when people use IPA or Variant Analysis and realize how much time it saves them, they're fine with it. They're always back for more."

And Hubank no longer has to worry about working with obsolete data. The Ingenuity Knowledge Base, the deeply integrated database at the heart of all the Ingenuity analysis tools, is manually curated by a team of expert analysts to ensure that it incorporates the most recent information. "I've got confidence that it's being maintained and kept up to date," Hubank says. "It's worth paying for that."

TRUSTED DATA

Ingenuity's proprietary database was a key factor in Hubank's decision to use both IPA and Variant Analysis. "The Knowledge Base matters because it gives you extra content that you're not getting from the public databases," he says. The Knowledge Base integrates detailed biological findings from the peer-review literature with related data from sources such as RefSeq, OMIM, and ClinVar as well as information on drug compounds and much more. The repository, which has been manually curated and continually improved for more than a decade, offers scientists the ability to find relevant biological connections faster and more reliably than would otherwise be possible.

For example, a scientist trying to link a particular gene to a certain phenotype might be confused by finding out that the gene is involved in a pathway

that doesn't seem related. "It doesn't make very much sense until you look it up in IPA, and all of a sudden you see the pathway is connected with a whole bunch of things in an area that you're not familiar with," Hubank says. "You realize that maybe the mechanism is through that area rather than the area you already know."

At UCL Genomics, Variant Analysis is most commonly used in the search for the causative variants behind rare diseases. Hubank made the choice to go with the application because it lets him empower individual users to make their own decisions about the variant search. "We can generate calls in our core lab, but it's not always appropriate for us to do so because we don't necessarily know the clinical context of the samples," he says. "It's better if the user is working within Variant Analysis to generate the calls and interpret their importance." For example, Hubank's partners at the Centre for Translational Medicine handle most of the gene discovery for projects from Great Ormond Street Hospital.

His team also relies on IPA, Ingenuity's pathway and network analysis tool for 'omics data, to track gene function and relationship to other genes. "It's extremely useful in terms of speeding things up," he says. "It's also able to identify connections you might not expect or wouldn't have thought of."

They have also found value in using IPA and Variant Analysis together. Scientists studying a rare genetic disease, for example, could run their list of candidate genes through Variant Analysis to get useful information on which candidate variants are most likely to be causative. Armed with that list, they could then gain an even deeper insight by turning to IPA, which serves as a one-stop shop for learning about those high-ranking candidate variants. "IPA gets you up to speed with a list of everything that's known about those candidates so that you're then in a better position to make a decision about where to go next," Hubank explains. What users emerge with is not just data, but actionable knowledge.

Ingenuity's web-based applications don't replace the need for smart and dedicated scientists who are passionate about their research. "It doesn't do all the work for you," Hubank says. "But it puts you in a better place to start attacking that problem."

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

