



Genomenon Using NIH Grant to Add AI to Variant Interpretation Platform

Aug 08, 2018 | [Neil Versel](#)

CHICAGO (GenomeWeb) – After a successful proof-of-concept study in 2017, Genomenon is well into its work under a two-year, \$1.5 million National Institutes of Health grant to add artificial intelligence to variant interpretation.

Ann Arbor, Michigan-based Genomenon is automating the curation of literature through a technique it calls [genomic language processing](#). The company said that the NIH grant will allow it to apply GLP to American College of Medical Genetics and Genomics-Association of Molecular Pathologists guidelines for determining the pathogenicity of genetic variants.

NIH's National Human Genome Research Institute [published](#) the Phase II Small Business Innovation Research award in January, but just recently delivered the money, according to Mark Kiel, Genomenon cofounder, CSO and vice president of product strategy.

Phase I, a six-month proof-of-concept study, was [awarded in early 2017](#).

"We set goals for ourselves for the sensitivity and the specificity of the database that we're assembling, the variant database from the medical literature, comparing to manually curated gold-standard databases that existed at the time," Kiel said of the preliminary phase. "We exceeded those goals that we set out for ourselves," he added, though Genomenon has not released any statistics yet.

The new phase is about bringing machine learning and artificial intelligence to the data compendium from Phase I. It relies on Genomenon's core [Mastermind](#) technology, an analytic and data visualization tool that mines millions of medical publications to deliver a list of disease-gene-variant relationships curated from primary medical literature, prioritized by strength of evidence.

Mastermind aggregates variant data from medical literature, then returns results based on disease indication, patient loss of function, or biological consequences of variants. "We show all the papers where the variants were published and we show the scientific context in which they were described," Kiel said.

"Aggregation, organization, and then evidence presentation, that is what Mastermind is, principally, and that is what we worked on for Phase I," he continued. "Now, as we're moving into Phase II, we don't want to rest on our laurels. We want to make good use of that information and knowledge that we've organized, and we want to even further facilitate automation of variant interpretation."

To get to that point, Genomenon needs to train its algorithms with artificial intelligence and improve the way results are displayed.

"The way that we're doing that in Phase II is by better prioritizing which pieces of evidence — which papers or which paragraphs — are the most relevant to a specific clinical concern,

such as ACMG classification guidelines when looking for functional studies of consequence for a particular missense variant in one gene," Kiel said.

As with any AI, the more data collected, the more accurate the predictions can be. Kiel mentioned the importance of sensitivity and specificity.

"I need to be sure that my search is maximally sensitive, and that comes out where we talk about how comprehensive Mastermind's reach into the literature is," he said.

"Then there's also specificity. We don't want to look at things that are off target. You could have a maximally sensitive database that you would spend three decades combing through. That is not a viable solution because the results lack specificity," Kiel added.

"Striking a balance between sensitivity and specificity, where you're not unduly sacrificing either, that is the real Holy Grail of automating data analysis. We're using our information, the data substrate from the medical literature, to drive our machine learning algorithms to produce the most sensitive and specific information to our users."

This is new ground for bioinformatics, which is why Genomenon calls its technology genomic language processing, rather than natural language processing. The company said in a press release that there is "nothing 'natural' about the dozens of different ways that authors may describe genetic variations in the scientific literature," so algorithms have to be trained to spot multiple keywords for the same concept. Otherwise, curation requires manual intervention.

"We're taking those manual approaches and systematizing them using our data within our platform," Kiel said, describing the task as "extremely challenging."

SBIR grants often go to early-stage companies, but three-year-old Genomenon has already hit the commercial market with Mastermind. The company offers both a free and a paid "professional" version. The latter can handle higher-throughput, automated data analytics, but both editions contain the same datasets.

"We're not pulling any punches in the free version. We want the community to benefit from the work that we've done and we want those who make maximal use of what we've done to help support the business activities that Genomenon is pursuing," Kiel said.

In the next few weeks, Genomenon plans on making its data freely available in VCF files for easier integration into clinical information systems. "We're releasing the information to a broader spectrum of users, both commercial and academic, in the hope that there will be value that's generated above and beyond what we could generate ourselves," Kiel said.

Since work began on Phase II of the NIH grant, Genomenon has made its technology capable of identifying variant references in the literature that are "most appropriate to the ACMG-AMP classification schema," Kiel said. "Finding those references, prioritizing which of those are the most meaningful, and pre-organizing or pre-assembling what information in those papers might be meaningful for variant interpretation, those are all capabilities that we have already."

Kiel said the company also is working with its partners, including Veritas Genomics, [LifeOmic](#), [Saphetor](#), and other, undisclosed entities, to make its technology more friendly for clinical environments and to demonstrate the efficacy and accuracy of the Mastermind pipeline. Kiel also said to expect publication of a manuscript soon on some of the work with Veritas.

Those partners and their clients are helping Genomenon to expand the scope of Mastermind use. "We have serendipitously come across users who are doing more large-

scale discovery work, primarily in pharmaceutical arenas," he said. The company also has worked with reference and commercial laboratories on biomarker discovery and evidence organization.

"Mastermind is intended to be searched one variant at a time, but we have at our disposal this great aggregation of all of this genomic content. If you take a step back and look at the forest for the trees, you are able to see patterns that ... only can emerge when all of this data is assembled," Kiel said.

"Having at our disposal all of this unbiased data from the whole complement of the medical literature, that allows us to see very broad patterns with all of the detailed, specific information to back up the claims that we're making about this biomarker being relevant to this drug indication or this disease pathogenicity," Kiel said.