



## Genedata Moves into Next-Gen Sequence Analysis Market with New Expressionist Module

October 01, 2010

By [Uduak Grace Thomas](#)

**This week**, Genedata threw its hat in the next-generation sequence analysis market with the release of a new module for its enterprise Expressionist analysis platform targeted at NGS data.

The Genedata Expressionist Refiner Genome module is designed to process and analyze next-gen sequence data and adds to the Refiner Array and Refiner MS Expressionist modules that the company already offers for microarray and mass spectrometry analysis, respectively.

The company was founded in 1998 with an initial focus on the microarray market, and the new module marks its first product for the sequencing sector.

Jens Hoefkens, head of the Genedata Expressionist business unit, told *BioInform* that the company aims to address the demand for better analytical solutions to handle large amounts of sequence data. "With every new profiling technology the amount of data that gets generated grows exponentially," he said. "With MS, it is not unusual to have a single experiment consist of 100 gigabytes of raw data and with NGS it is not uncommon to have a terabyte of data from a single experiment."

Given this proliferation of information, Genedata added the Refiner Genome module to its Expressionist suite of tools to enable users to visualize terabytes of data and perform whole-genome transcriptome, methylome, gene regulation, copy number variation, and SNP analysis.

Hoefkens explained that Refiner Genome takes as input any data that can be mapped to a reference genome, such as copy number variation data, methylation data, and gene expression data. It then reads the data and converts it into a format that is "more manageable in size." As a next step, Hoefkens continued, the module preps the data for statistical analysis by assigning numbers to the data — for example, one number per gene or one number per exon — which makes it easier, for example, to identify specific genes that are linked to cancer.

Genedata collaborated with the genomic analysis laboratory at the Salk Institute to develop the Refiner Genome module. The lab is currently using the software to analyze DNA methylation, RNA sequencing, and ChIP sequencing data and is providing feedback to the company.

Joseph Ecker, who heads Salk's genomic analysis lab, said that the module lets researchers "bring together all these data types on a single platform that includes intuitive visualization capabilities," which will help researchers "use experimental data in their research."

Bob Schmitz, a research associate in Ecker's lab at Salk, told *BioInform* that the lab opted for the module primarily for its ability to create easy-to-use workflows and to integrate multiple data types. The lab is using the software to profile the epigenomes of strains of *Arabidopsis thaliana* and human cell samples, among other projects.

He explained that most researchers at his lab are wet bench scientists with little computational experience. These colleagues have no problem designing experiments and working with "a gauntlet of different sequencers and applications," but run into trouble when they try to analyze the terabytes of data generated in their experiments.

"That's really hard to do for those of us who aren't trained as computational biologists," Schmitz said. With Refiner Genome, he said, "we are able to bring in all these datasets and visualize them all in one browser and then you are able to play around with setting different thresholds or browsing through your data."

Furthermore, Schmitz said, new researchers who join the lab can use the module for their data analysis, which saves time that would have otherwise been wasted training each researcher "individually in the command-line environment."

He added that Refiner Genome replaced pipelines that Salk researchers had created to analyze their DNA methylation data in the early days of whole-genome methylation analysis research.

Hoefkens cited Refiner Genome's ability to integrate hundreds and thousands of data points generated on different types of platforms, as well as its ability to let users visualize and interact with their raw data, as key features.

The goal of most large-scale experiments is to find the "gene that makes the difference" Hoefkens noted, adding that Refiner Genome "takes away a lot of the hay from the haystack so ... it becomes much easier to find the needle."

The module lets users select and order the processing steps in their workflows and then change the settings by double-clicking on each individual step.

Once workflows are created, users can save them and share them with other researchers via e-mail or send them within the application. This way, the workflow "scripts" can be

reused for future experiments that use the same types of data, Hoefkens explained. For example, a workflow that's designed to process RNA-seq data can be reused multiple times rather than creating a new workflow each time.

Schmitz noted that this feature makes working with other labs easier because users can create a workflow or pipeline, "lock it up," and then share it with other researchers.

Also included in the module is a genome browser for visualizing raw data. Users can zoom in on specific regions of a chromosome to look for gene expression, for example, and use the tool to create an "all-in-one plot" that shows gene expression data from the same region of the chromosome in multiple samples. Users can view separate plots of genes of interest to see which are over-expressed, for example.

The Expressionist platform also includes the Genedata Analyst module, which lets users perform statistical analysis on their condensed data like box-plot and fold-change analyses as well as T-tests. Hoefkens said that the Analyst module is separate from Refiner because it makes it easier to integrate data from multiple platforms and it also gives users the flexibility to purchase only the modules they need.

Once the statistical analysis is complete and genes of interest have been identified, users can export their data back into Refiner Genome for further analysis.

While Genedata faces considerable competition in the next-gen sequence analysis market, Hoefkens said that while many companies in the sector focus primarily on developing tools for mapping sequences, Genedata has "deliberately stayed away" from that space, choosing instead to focus on its "strengths" in "numbers and the visualization of raw data."

He noted that software vendors with tools for mapping sequences are "potential partners" for Genedata and that the company is considering partnerships with some of these groups, but he declined to provide further details.

Going forward, Hoefkens said the company plans to focus primarily on developing tools in the Refiner module for SNP management and data analysis as well as other "individual activities and small improvements." He noted that the company plans to continue collecting feedback from customers and incorporating some ideas into future incarnations of the module.