

The MDL-Partek Alliance: Powering Decision Support with a Robust Statistical Engine

MDL has raised the stakes in the handling of massive discovery data sets by signing a three-year marketing agreement with Partek Incorporated, a St. Louis-based provider of statistical analysis and pattern recognition software used in the life sciences and engineering industries for a variety of applications including the analysis of genomic and proteomic data, high-throughput screening, and computer-aided drug design.

The agreement makes MDL the authorized exclusive, worldwide marketer of Partek's data analysis and visualization products for the life sciences market. The products covered by the agreement are:

- Partek Pro, the company's flagship data processing, analysis, and visualization system
- Partek Discover, a toolkit for exploratory data analysis
- Partek Infer, used for statistical inference
- Partek Predict, used for predictive modeling
- Partek Connect, used to integrate Partek software products with database, Web, and third-party software applications

Tom Downey, president and CEO of Partek, recently introduced his company's products to MDL's sales and marketing staff. We sat down with Tom to discuss his research philosophy in greater detail and to explore how Partek's powerful statistical tools complement MDL's extensive offerings in discovery informatics.

MDL: *What is your philosophy when it comes to turning data into discovery?*

Tom Downey: Turning data into discovery involves a multi-step, methodical approach to data analysis that begins with data inspection. During the data inspection step, we use exploratory and statistical techniques to look for anomalies in the data. In the next step, we take an open-ended, data-mining approach and ask, are there patterns in the data set? Eventually, we reach the point where we can formulate hypotheses, meaning we start asking more specific questions, such as, are any of these molecular descriptors well correlated with response activity, or do any of these genes differentiate different cell types—that type of thing.

MDL: *How do Partek products exploit the synergies between visualization and statistics?*

Tom Downey: Partek's visualization tools are geared towards the exploratory analysis of high-dimensional, multivariate data.

Taking advantage of statistics early in the research pipeline, Partek products create powerful visualizations using statistical methods such as principal components analysis and multidimensional scaling. Our products also apply statistical methods to finding "needles in the haystack" and then present these statistical findings through compelling visualizations.

Either way you look at it, the statistics and the visualizations are thoroughly integrated. With Partek's tools linking visualizations to statistical results, researchers don't have to

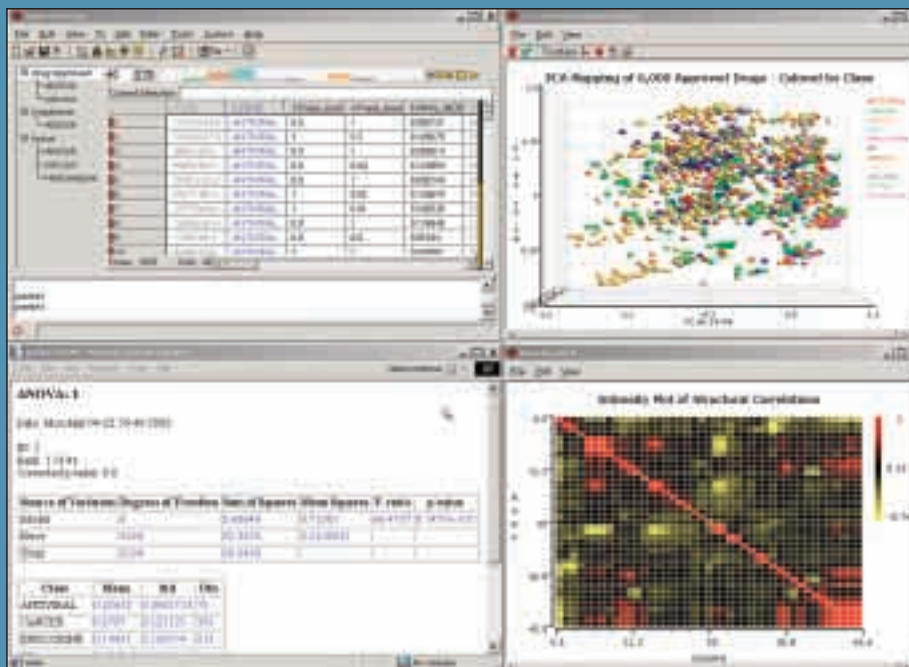
fish through mounds of data looking for interesting patterns. The patterns are linked to the statistics that created them in the first place.

MDL: *How much of a statistical background must a researcher have to use Partek products effectively?*

Tom Downey: Any researcher who works with a lot of data should know something about statistics. However, I like the analogy of a car company that makes fine motorcars and markets its products on the exemplary engineering that goes into every car. In the automotive world, you don't need to be a mechanical engineer to appreciate and gain maximum value from a well-built, finely tuned car. The same is true with statistics and data. Researchers don't need to be expert statisticians; they just need to know when and how to use the right tools.

MDL: *Tell us a little bit about the typical users of Partek products.*

Tom Downey: Partek products are best suited for answering questions in the life sciences, where so many underlying mechanisms are unknown and large amounts of empirical data—millions of rows or columns of numbers—need to be analyzed. Typical users are researchers who want to understand what allows a drug to be absorbed into the system, what makes a drug toxic, or how certain genes work together.



Partek-Integrated Visual and Statistical Analysis Capabilities

The images illustrate the visual and statistical analysis of 6,000 approved drugs. In the upper-right region, the drugs are labeled and colored as belonging to one of 11 classes: Antiviral, Cancer, Cardio, CNS, Endocrine, GI, Immuno, Infect, Inflamm, Metab, or Pulmonary. Structural properties for each drug are stored in the Partek Pro Analytical Spreadsheet® shown in the upper-left region. Principal components analysis maps the 43 structural properties to three dimensions for visualization. Each point in the scatterplot corresponds to an approved drug and is colored by its class. The lower-right region is an intensity plot showing structural correlations. The lower-left region is a browser containing the results of an analysis of variance (ANOVA) statistical analysis.

MDL: *How are Partek products assisting in discovery research today?*

Tom Downey: We're working with the National Institutes of Health, looking primarily at gene expression data and helping to identify genes that differentiate tissue or cell types. Working with a data set that included 12,000 variables and taking advantage of Partek's ability to integrate statistics with visualization, we discovered anomalies that led the NIH researcher to realize that test tubes had been mislabeled. Any time we can help life science researchers make better decisions, based on the way things really are, we feel our products have lived up to their reputation.

MDL: *To what do you attribute Partek's success to date?*

Tom Downey: I can think of two things immediately. One is our focus on providing the most powerful statistical analysis and visualization software to help scientists get answers, solve problems, and make discoveries. And then there's our equally strong focus on customer support—making sure customers get maximum value from their Partek toolkits.

For example, we'll be working with MDL to develop interfaces between the Partek products covered by our agreement and MDL's ISIS and Assay Explorer products. Our joint goal is to make sure that researchers have the benefit of powerful, integrated statistical tools and the support required to use the tools effectively.

MDL: *How do you see Partek's products interacting with and complementing MDL's products?*

Tom Downey: There are many products out there that handle data visualization and many products that do statistical analysis, but Partek is the only company that integrates these two critical capabilities. The MDL-Partek alliance is taking this product offering to the next level by tightly meshing these capabilities with MDL's comprehensive experiment management and decision support solutions.

Ultimately, I envision massive amounts of customer data flowing seamlessly between Partek's statistical engines and MDL's experiment management and decision support solutions, with new discoveries continuously flowing back into the research pipeline for use in ongoing discovery projects. ■