

> Mining data helps save children with brain tumors

Dr. Eric Bremer turns to bioinformatics to solve pediatric brain tumors

Each year, nearly 3,000 children in the U.S. are diagnosed with brain tumors. Almost half will die within five years, making it the most fatal cancer among children. If a child does survive a brain tumor, the long-term effects can be significant and include neurological disabilities, retardation, and psychological problems.

Beyond surgery, successful treatments for pediatric brain tumors are rare. In fact, clinicians must often conduct experimental research and protocols to develop the treatments. Dr. Eric Bremer, a former director of brain tumor research at Children's Memorial Hospital in Chicago, is regarded as one of the leading scientists searching for a better way to treat brain tumors. One of his main goals is to build a gene expression database for pediatric brain tumors, and to then correlate this with both past and ongoing research on effective treatments.

Challenge

One of the first requisites for this cancer treatment database is accurately classifying the tumor. While the vast majority of pediatric brain tumors can be categorized as gliomas, ependymomas, and medulloblastomas, altogether there are 12 or so tumor types, including subtypes. Classification is very subjective because pediatric brain tumors are often difficult to distinguish by appearance, and there are few objective markers such as those found for other childhood cancers such as leukemia.

At a glance

Country: USA

Industry: Healthcare

Founded: 1882

Company type: Private

Revenues: \$240.2 million

Employees: 1,100+

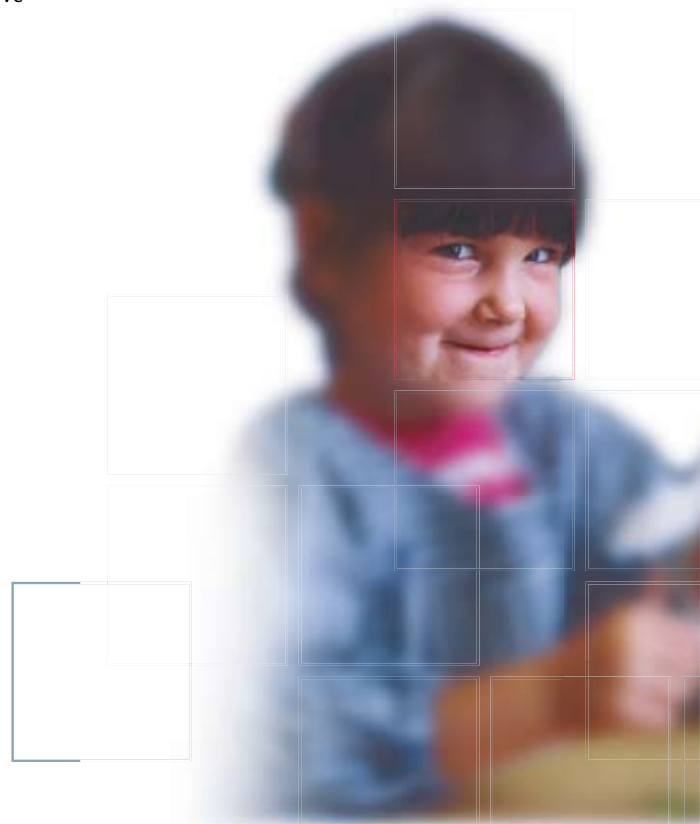
Application

Bioinformatics

Solutions used

PASW® Modeler*

*PASW Modeler, formerly called Clementine®, is part of SPSS Inc.'s Predictive Analytics Software portfolio.



In addition to tumor type, it's important to classify a tumor's stage or grade. Cancers are generally stratified into four stages, from the most benign (stage I) to malignant (stage IV). Neuropathologists often find it difficult to distinguish between intermediate stages. However, the treatment for two different stages can be drastically different, and incorrect staging can have dramatic consequences for the patient. For example, if a child with a stage two tumor is misdiagnosed with a more advanced grade, he would unnecessarily receive more aggressive treatment. This not only results in needless pain, but also could lead to long-term damage.

Solution

A key to accurate classification of pediatric brain tumors lies at the molecular level. Just as a skin cell and a liver cell vary in their gene expression patterns, the same is true for different tumors or tumor grades. Dr. Bremer can capture these differences with gene expression microarray experiments—but analyzing these differences to identify and classify the tumor is a more formidable obstacle. Said Dr. Bremer, “You can easily get 7,000 to 30,000 data points for each sample. The problem is how to make sense of them.”

Dr. Bremer literally ran into the solution to his problem at the Microarray Data Analysis Conference, where he met two SPSS Inc. employees. Later, they consulted and worked with Dr. Bremer to analyze his microarray data using SPSS Inc.'s data mining solution, PASW Modeler, the leading data mining workbench. Traditionally used in customer-oriented business intelligence applications, PASW Modeler is now being used in the life sciences to study microarray data.

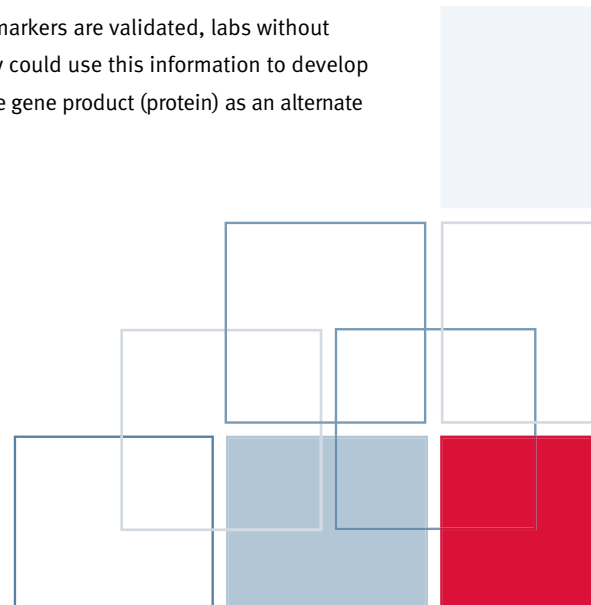
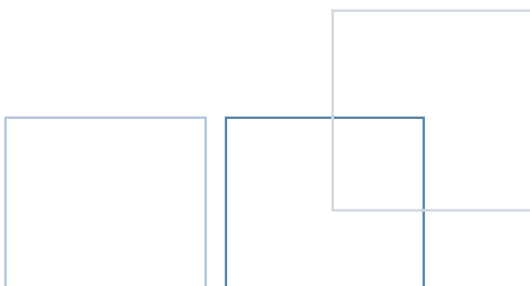
The power of PASW Modeler

Dr. Bremer was won over by PASW Modeler for one simple reason: “it worked.” Dr. Bremer combined his own data with that of the publicly available Pomeroy et. al. data set (*Nature* vol. 415, pp. 436-442, 2002) resulting in a total of 133 tumor samples from the six major pediatric brain tumor types. This software successfully classified these tumors with greater than 95 percent accuracy. While these samples were well described pathologically, they served as a test case that bodes well for future pediatric brain tumor classification, especially difficult-to-classify tumors. In addition, sub-classification of gliomas and medulloblastomas also appeared to be very reliable.

The key was the software's flexibility. Prior to using PASW Modeler, Dr. Bremer used up to five different software packages to analyze microarray gene expression data. This software replaced all of those packages.

“There is no one right way to analyze the data; you want to try several ways,” said Dr Bremer. “What's so great about PASW Modeler is that it's a workbench. It's easy to try different methodologies in order to find a method, or combination of methods that works best.”

Dr. Bremer used two of the software's predictive models, artificial neural networks and decision trees, to analyze and classify his data. Information from one model complemented the other. While the neural network resulted in a more accurate classification, it didn't show how it actually accomplished the classification. The decision tree, however, showed precisely how the tumors were classified and revealed potential gene markers that characterized certain cancers. Once these markers are validated, labs without microarray technology could use this information to develop antibodies against the gene product (protein) as an alternate means of diagnosis.



Just as important as the model is the data that the model is based on. Microarray data analysis presents a number of challenges given the small number of samples and large number of genes. Dr. Bremer's data set had 133 samples, but nearly 7,000 variables.

It's also important that differences in gene expression values genuinely reflect biological variation and not artificial differences introduced during sample preparation. PASW Modeler can help assure this is the case by processing the data quality parameters along with the samples. This software also simplifies the task of feeding data into the model. Prior to PASW Modeler, Dr. Bremer had to organize the expression values in a specific format dependent on the requirements of the analysis package. Dr. Bremer now skips this step because PASW Modeler streams can automatically prepare and accept gene expression data directly from his database—a huge time saver.

More is better

Classification using data obtained from gene expression microarray experiments is just part of a bigger picture. Expression measures RNA levels and not the final gene products. Dr. Bremer's database will eventually incorporate clinical, pathological, and biochemical information to provide as complete and accurate a picture as possible. In addition, patient outcome information will be added so that the database may reveal which treatments work best with brain tumors sharing genetic and pathological characteristics. This is the point at which bioinformatics transition to "biomedical informatics."

Dr. Bremer also realizes the critical need to share data and knowledge. To this end, he plans to deploy PASW Modeler streams via a Web server so that any pediatric brain tumor researcher can submit files from his or her own microarray experiments for analysis, and then receive predictive responses. PASW Modeler will enable Dr. Bremer's data to be combined with data from other researchers. In turn, the added data can be used to update and add value to the original data set; the more samples his database contains, the more accurate it will be and the more lives that will be saved.

The bigger database will also help reveal the multitude of genes that are involved in tumor development. According to *Nature*, in 2001 alone, more than 21,000 articles on characterizing, diagnosing, and treating malignancies were published. Dr. Bremer will use text analytics software from SPSS Inc. to sift through this literature and extract patterns that, for example, when combined with data from his tumor database, might help him evaluate prime drug targets that would form the basis for a cure for cancer. This, after all, is Dr. Bremer's ultimate goal.

□ "There is no one right way to analyze the data; you want to try several ways," said Dr. Bremer. "What's so great about PASW Modeler is that it's a workbench. It's easy to try different methodologies in order to find a method or combination of methods that works best."

— Dr. Eric Bremer
Director of Brain Tumor Researcher

For more customer stories, visit www.spss.com/success.

SPSS is a registered trademark and the other SPSS Inc. products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2003 SPSS Inc. All rights reserved. MDCCS-0209

