

Genedata vereinfacht Suche nach Biomarkern

Biomarker spielen eine Schlüsselrolle, wenn es darum geht, zum Beispiel die Wirksamkeit beziehungsweise Nebenwirkungen eines neuen Arzneimittels zu belegen. Damit dies zuverlässig und in kürzester Zeit gelingt, werden mittels modernster Labortechnologien riesige Mengen hochkomplexer molekularer Daten generiert. Anschließend müssen diese Daten effizient analysiert werden. Hierfür hat das in Konstanz und Basel ansässige Unternehmen Genedata das Software-System Genedata Expressionist® entwickelt, welches basierend auf biostatistischen Methoden das Zusammenwirken von Genen, Proteinen und Metaboliten aufschlüsselt und sie als Biomarker qualifiziert. Ein Großteil der weltweit führenden Pharmafirmen setzt die Software mittlerweile ein.

Das Hauptproblem bei der Suche nach Biomarkern ist, diejenigen Gene, Proteine und Metaboliten aus zigtausenden von potenziellen Molekülen zu identifizieren, deren Quantitäten sich zum Beispiel als Indikator für die Wirksamkeit eines Medikaments eignen. Um die große Anzahl biochemischer Moleküle aus einer Blut- oder Gewebeprobe zu quantifizieren, sind für jeden Molekültyp spezielle Messtechnologien notwendig. Diese in den letzten Jahren entwickelten Technologien können zum einen zehntausende solcher Moleküle pro Probe gleichzeitig messen, zum anderen im Hochdurchsatz tausende von Proben in kürzester Zeit prozessieren, so dass für ein bestimmtes wissenschaftliches Projekt schnell Terabytes an experimentellen Rohdaten anfallen, die es zu analysieren und interpretieren gilt.

Genau das leistet die vom Unternehmen Genedata entwickelte Anwendung Expressionist® außerordentlich zuverlässig und schnell. „Es können Terabytes von Rohdaten unter Anwendung innovativer mathematischer und statistischer Methoden eingelesen und prozessiert werden“, erklärt Dr. Timo Wittenberger, Leiter der Genedata-Niederlassung in Konstanz. Der Output ist zum Beispiel eine limitierte Liste von Genen, Proteinen und Metaboliten, die als Biomarker verwendet werden können, um entweder die Wirkung von Medikamenten besser vorhersagen zu können, oder zum Beispiel Patienten zu identifizieren, für die ein bestimmtes Medikament gut wirksam ist.

„In einer initialen Phase integriert die Software Daten aus verschiedenen experimentellen Prozessen in einen gemeinsamen Datensatz, überprüft die Datenqualität, identifiziert experimentelle Fehler und liefert einen Qualitätsreport“, erklärt Dr. Timo Wittenberger. Wichtig hierbei ist, sicherzustellen, dass ein qualitativ hochwertiger Datensatz frei von jeglichen experimentellen Artefakten (z.B. durch Instrumentenfehler) für die nächste Phase der Analyse zur Verfügung steht. Hierbei spielen auch visuelle Methoden eine wichtige Rolle, die es dem Wissenschaftler erlauben, schnell und möglichst intuitiv aus mehreren hundert Gigabyte von



Dr. Timo Wittenberger, Leiter der Genedata-Niederlassung in Konstanz, an seinem Arbeitsplatz
© privat

Daten potenzielle Fehler zu erkennen.

Fragestellung setzt unterschiedliche Module in Gang

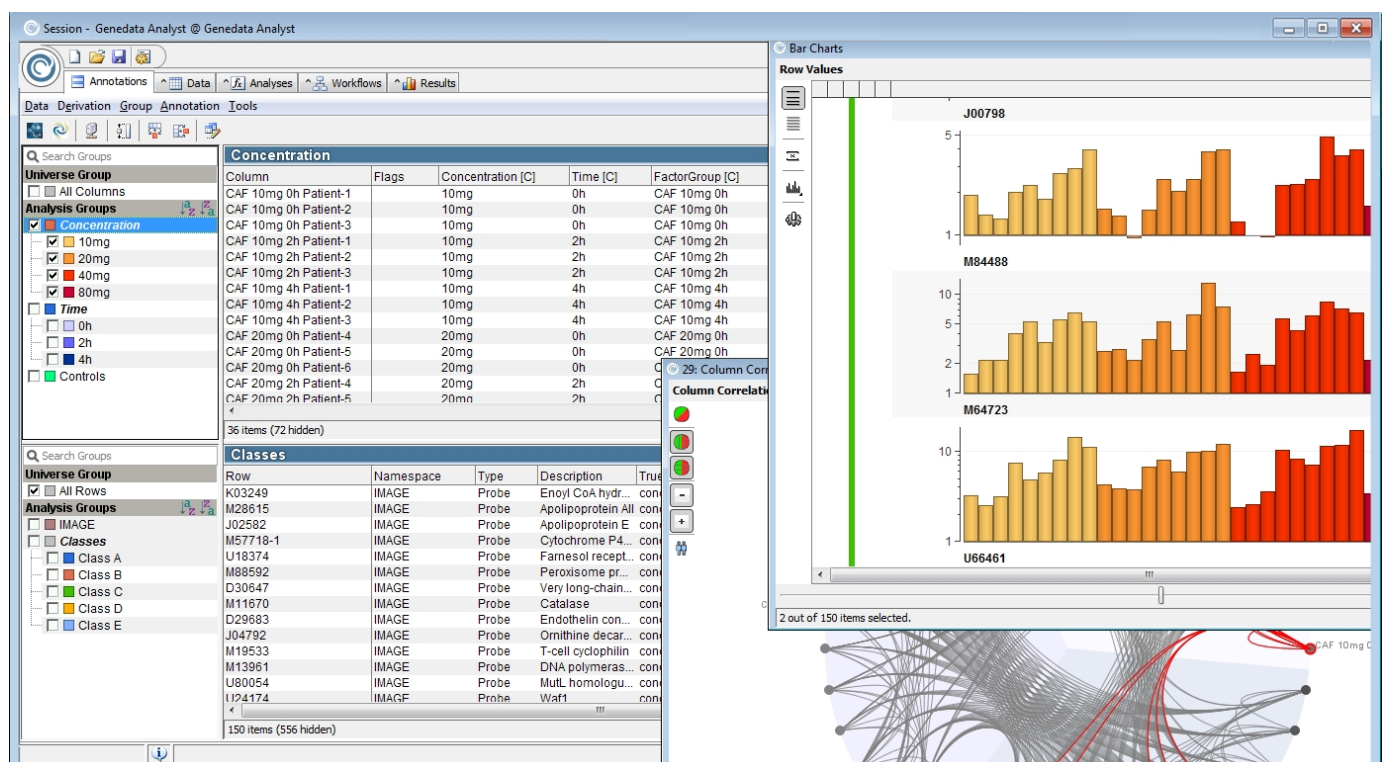
Der modulare Aufbau der Software erlaubt es den Wissenschaftlern, technologiespezifische Module einzusetzen, um bestimmten Forschungsumgebungen gerecht zu werden. Geht es etwa um die Analyse von Protein- oder Metabolitenzusammensetzung biologischer Proben basierend auf Massenspektrometrie, kommt zum Beispiel das Modul „Refiner MS“ zum Einsatz. „In diesen Profilierungsstudien werden bei der Suche nach bisher unbekanntem Biomarkern in einem einzigen

Experiment tausende von Massenspektren aufgenommen“, so Dr. Wittenberger. Nachdem die Bestandteile der Probe mittels eines Chromatografieverfahrens aufgetrennt und durch Ionisation bzw. Fragmentierung molekular analysiert worden sind, generiert das „Refiner MS“-Modul aus den tausenden Massenspektren dreidimensionale Datensätze, die sich aus den drei Größen Laufzeit, Masse-Ladungs-Verhältnis und Intensität zusammensetzen. „Die Spektren müssen vom Rauschen befreit und genau aneinander ausgerichtet werden, bevor man sie mit anderen Experimenten vergleichen kann“, erklärt Dr. Wittenberger. In diesen Spektren erzeugt jedes Protein bzw. jeder Metabolit eine Reihe von Signalen, die die Wissenschaftler über das „Refiner MS“-Modul und mit Hilfe von im Internet verfügbaren Repositorien von MS-Spektren identifizieren können. Die Software selbst verfügt neben den Datenanalysetools auch über eine Datenbank zur Speicherung, Verwaltung und Suche von Biomarker-Daten.

Rekonstruktion von Stoffwechselkreisläufen

Bei der Bestimmung von Biomarkern aus der großen Anzahl von quantifizierten Molekülen kommt das Analyse-Modul „Genedata Analyst“ ins Spiel, das die Daten einer integrierten Analyse zugänglich macht. „Ein großer Teil dieser Daten ist nichts weiter als natürliches 'Rauschen', das von bedeutsamen Signalen unterschieden werden muss“, so Dr. Wittenberger. Tatsächlich stellen solche Daten eine große Herausforderung an den Wissenschaftler, denn es werden für eine relativ kleine Anzahl an Experimenten sehr große Datenmengen erzeugt. Nur durch statistische Methoden ist es ihm möglich, das Signal-zu-Rausch-Verhältnis und die Signifikanz einzelner Ereignisse zu bestimmen. Dabei wird auch auf Klassifizierungs- und Gruppierungsverfahren zurückgegriffen.

„Der große Vorteil des Analyst-Moduls ist, dass dieses selbst bei großen Datenmengen mit über einer Milliarde Datenpunkten einsetzbar ist“, so Wittenberger. Außerdem hilft das Tool dabei, Daten aus unterschiedlichen experimentellen Domänen zusammenzuführen. So ist es beispielsweise möglich, Metaboliten und Proteinexpression direkt miteinander zu vergleichen und Stoffwechselkreisläufe zu rekonstruieren.



"Genedata Analyst" hilft, das Signal-Rausch-Verhältnis aus der großen Anzahl quantifizierter Moleküle zu bestimmen und somit die Suche nach Biomarkern zu vereinfachen.

© Genedata

Modulare Vielseitigkeit unterstützt Chip-Technologien

Bei der Suche nach Biomarkern kommen weitere Elemente der Genedata-Technologie ins Spiel: „Mit dem Modul „Refiner Array“ können tausende von Microarrays gleichzeitig normalisiert werden“, beschreibt Dr. Wittenberger die Funktion. Durch Normalisierungsverfahren werden in der Regel technisch bedingte Schwankungen in den Messwerten ausgeglichen. „Eine einfache Form der Normalisierung ist das Skalieren, bei dem alle Messwerte eines Experiments mit einem bestimmten Faktor multipliziert werden, so dass der Mittelwert aller Messwerte zum Beispiel gleich 1.000 ist“, erklärt Timo Wittenberger. Bei der Analyse des Transkriptoms, also der Summe aller in einer Zelle hergestellten RNA-Moleküle, wird DNA-Chip-Technologie eingesetzt, die die Aktivität bekannter Gene auf einem Microarray mittels Hybridisierung misst. Das Modul „Refiner Array“ nimmt die Rohdaten aus Chip-Experimenten als Grundlage und führt zunächst Qualitätskontrolle, Normalisierung und Kondensierung durch.

Bei Daten, die mit den neuesten Hochdurchsatz-Sequenzieretechnologien erzeugt wurden, übernimmt das Modul „Refiner Genome“, da diese meist im Kontext des Genoms analysiert werden. „Das Modul ist nicht nur in der Lage, große Datenmengen aus Hochdurchsatz-Sequenzierern zu prozessieren, sondern es kann auch Daten aus Chip-Experimenten visualisieren, damit beide Datentypen direkt im genomischen Kontext dargestellt werden können“, erklärt Dr. Wittenberger. Dies können Expressionsdaten sein, sofern die chromosomale Lokalisation der entsprechenden Gene bekannt ist, oder Daten zu Polimorphismen (SNPs), DNA-Methylierung und Copy Number Variations (CNVs), die alle mit DNA-Chips gemessen werden können.

Unterstützung bei Profiling- und Kinetikexperimenten

Ein anderes Beispiel für die Anwendung von Genedatas Software ist die Analyse der Kinetik einer Substanzwirkung. Hier werden häufig über einen definierten Zeitraum experimentelle Daten erfasst, und diese zeitliche Korrelation muss in der Analyse möglichst effizient eingerechnet werden. „Besonders in Vorexperimenten solcher Studien sind die statistischen Methoden von großer Bedeutung, da sie Aussagen über das optimale Design der Hauptstudie geben“, so Timo Wittenberger. Da in Kinetikexperimenten besonders große Datenmengen anfallen, ist eine Hochdurchsatz-Software wie „Genedata Expressionist“ für die effiziente Prozessierung und Analyse notwendig. Das trägt dazu bei, dass Medikamentenexperimente effizienter und schneller durchgeführt werden können, was letztendlich dem Endverbraucher zugute kommt.

Anwendung auch in der Pflanzenbiotechnologie

Neben der pharmazeutischen Forschung kommt die Hightech-Software zum Beispiel auch in der Pflanzenbiotechnologie zum Einsatz, wenn es um die Untersuchung von tausenden von Pflanzensorten pro Jahr auf ein umfassendes Set von Kriterien geht. „Ein flexibles und skalierbares System für die schnelle Analyse großer Mengen von molekularen Profiling-Daten ist essenziell für die Identifizierung und Charakterisierung der am besten geeigneten Kandidaten für die weitere Entwicklung und Zucht“, erklärt Dr. Wittenberger.

Mit der kontinuierlichen Weiterentwicklung seiner Software-Technologien wie Expressionist® hat sich Genedata in der Zwischenzeit eine weltweit führende Position erarbeitet. „Unsere Anwendungen kommen aktuell bei 22 der Top 25 pharmazeutischen Pharmafirmen weltweit zum Einsatz“, verrät Dr. Timo Wittenberger.

Fachbeitrag

28.11.2010

Andreas Zangar

BioLAGO

© BIOPRO Baden-Württemberg GmbH

Weitere Informationen

Dr. Timo Wittenberger

Head of Business Operations Konstanz

Genedata GmbH

Byk-Gulden-Straße 2

78467 Konstanz

Tel.: +49-7531-209901

Fax: +49-7531-209890

E-Mail: [timo.wittenberger\(at\)genedata.com](mailto:timo.wittenberger@genedata.com)

