

Entwicklung und Implementierung eines Auswertungswerkzeuges für Matrix-CGH-Daten

Kreuz M², Rosolowski M¹, Berger H², Wessendorf S³, Schwaenen C³, Hasenclever D²

¹Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig, Deutschland

²Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Deutschland

³Abteilung Innere Medizin III, Medizinische Universitätsklinik und Poliklinik Ulm, Deutschland
markus.kreuz@imise.uni-leipzig.de

Einleitung und Fragestellung Matrix-CGH (Comparative Genomic Hybridization) ist ein Verfahren zur hochauflösenden Analyse numerischer Unterschiede in der Kopiezahl von Chromosomen oder Chromosomenabschnitten. Diese spielen bei der Entstehung und Progression vieler Krebserkrankungen eine entscheidende Rolle. Durch vergleichende Hybridisierung verschiedenfarbig markierter DNA-Proben kann man genomische Imbalancen zwischen den Genomen der zugehörigen Gewebeprobe, z. B. Tumor und normaler Referenz, detektieren. Im Gegensatz zur konventionellen CGH werden kurze DNA-Abschnitte auf einem Chip als Sonden verwendet. Auf einem Chip können mehrere tausend DNA-Sonden untergebracht werden. Die Auflösung und Sensitivität der Messung ist dadurch deutlich höher als bei konventioneller CGH. Aufgrund des höheren Datenaufkommens stellt Matrix-CGH neue Anforderungen an die Auswertung und Interpretation der gemessenen Daten. Ziel war es daher, ein Werkzeug zu entwickeln, das die automatische Auswertung von Matrix-CGH-Daten ermöglicht.

Material und Methoden Die Anforderungsanalyse ergab, dass ein entsprechendes Auswertungswerkzeug folgende Analyseschritte unterstützen sollte:

Überprüfung der Datenqualität:	Selektion von Klon- und Chipmesswerten mit hoher Streuung, vielen ungültigen Messungen sowie hohem systematischem Fehler (Bias)
Normalisierung:	Normalisierung der Chipmesswerte
Einzelchipanalyse:	
Segmentierung:	Segmentierung der Messwerte der Einzelchips und dadurch Reduzierung des Einflusses der Messwertstreuung
Klassifizierung:	Einteilung der ermittelten Segmente in die Klassen normale Kopiezahl, Zugewinn und Verlust
Heterogenitätsmaß:	Bestimmung der Heterogenität eines Chips, d. h. Quantifizierung der Aberrationen
Multichipanalyse:	
Selektion von rekurrenten Regionen:	Automatische Suche nach Genomregionen in denen im untersuchten Datensatz häufig gleichartige Aberrationen auftreten
Reduktion der Einzelchips auf rekurrente Regionen:	Für jeden Chip werden die Messwerte in den ermittelten rekurrenten Regionen validiert. Weist ein Chip in einer rekurrenten Region eine Aberration entsprechend der rekurrenten Region auf, so ist diese Region im Chip präsent (1). Ist dies nicht der Fall, so ist die Region nicht präsent (0). Jeder Chip kann dadurch auf einen Vektor aus Nullen und Einsen reduziert werden.
Integrierte Auswertung von Matrix-CGH- und Genexpressionsdaten:	Wird eine Probe mit Matrix-CGH und Genexpressionsanalyse untersucht, dann soll für jedes gemessene Gen der CGH-Status geschätzt werden. Ein Gen wird anschließend durch seinen Expressionswert und seinen CGH-Status repräsentiert. Dieses Wertepaar bildet die Grundlage für weitere Analysen.

Ergebnisse Es wurde ein Auswertungswerkzeug mit der Statistiksoftware R implementiert, welches die aufgeführten Analyseschritte unterstützt. Die entwickelten Methoden werden an einem Matrix-CGH-Datensatz (n=213 Chips) des Verbundprojektes "Molekulare Mechanismen bei malignen Lymphomen" (MMML) illustriert.