

Patrick René Warnat
Dr. sc. hum.

Studienübergreifende Analyse von Tumor-Genexpressionsprofilen unter besonderer Berücksichtigung des Neuroblastoms

Geboren am 25.11.1978 in Gießen
Diplom der Fachrichtung Medizinische Informatik am 25.11.2003

Promotionsfach: DKFZ (Deutsches Krebsforschungszentrum)
Doktorvater: Prof. Dr. rer. nat. Peter Lichter

In den letzten Jahren wurde die DNA-Microarray-Technologie in vielen Studien zur Erstellung von Genexpressionsprofilen eingesetzt, insbesondere im Bereich der Krebsforschung. Die steigende Anzahl an veröffentlichten Genexpressionsdatensätzen macht den Vergleich und die Auswertung von Genexpressionsdaten verschiedener Studien mit ähnlicher Fragestellung wünschenswert. Allerdings existiert bislang kein „Gold-Standard“ für DNA-Microarray Experimente zur Genexpressionsanalyse, da verschiedene technologische Plattformen und experimentelle Protokolle nebeneinander existieren. Häufig werden in unterschiedlichen Studien sehr ähnliche oder gleiche Fragestellungen mit jeweils verschiedenen Microarray-Plattformen untersucht. Da die mit unterschiedlichen Microarray-Plattformen erzeugten Genexpressionsdaten quantitativ nicht direkt vergleichbar sind, sind spezielle Ansätze und Methoden nötig, um die Daten unterschiedlicher Plattformen gemeinsam auszuwerten. Gegenstand dieser Arbeit ist deshalb die studien- und plattformübergreifende Analyse von DNA-Microarray-Daten.

Die beiden im Bereich der Krebsforschung am häufigsten bearbeiteten Aufgabenstellungen bei der Analyse von Genexpressionsprofilen sind die Detektion von Genen mit differentieller Expression zwischen verschiedenen Phänotypen sowie die Erstellung von prädiktiven Modellen zur Phänotyp-Klassifikation auf der Basis von Genexpressionsdaten. Zunächst wurde untersucht, ob Genexpressionsdaten verschiedener Microarray-Plattformen zur Erstellung von prädiktiven Modellen kombiniert werden können. Dazu wurde der Ansatz verfolgt, Phänotyp-Klassifikation auf der Basis von diskretisierten Genexpressionsdaten durchzuführen. Für die Evaluation der plattformübergreifenden Phänotyp-Klassifikation wurden die öffentlich verfügbaren Genexpressionsdaten sechs verschiedener Studien verwendet, wobei jeweils zwei Studien die gleiche Tumorart untersuchten (Tumoren der Prostata, Tumoren der Brust, sowie akute myeloische Leukämie). In jedem Paar an Studien, welche die gleiche Tumorart betrachteten, wurden in jeweils einer Studie cDNA-Microarrays und in der anderen Studie Oligonukleotid-Microarrays verwendet. In einer plattformübergreifenden Analyse wurden hohe Klassifikationsgenauigkeiten (> 85%) erreicht, wenn zur getrennten Generierung und Validierung von prädiktiven Modellen zur Phänotyp-Klassifikation jeweils Proben aus beiden Studien eines Paares verwendet wurden.

Neben der plattformübergreifenden Phänotyp-Klassifikation wurde eine Meta-Analyse-Methode zur Detektion von Genen mit differentieller Expression eingesetzt, um Genexpressionsprofile des Neuroblastoms studienübergreifend zu untersuchen. Neuroblastom-Patienten zeigen heterogene klinische Verläufe, von spontaner Regression zum Teil bereits metastasierter Tumoren bis zur unaufhaltsamen Progression der Erkrankung. In neueren Studien wurden die Genexpressionsprofile von Neuroblastom-Tumoren mit verschiedenen klinischen Phänotypen assoziiert. Dennoch sind bislang für wichtige Patientengruppen mit dieser Erkrankung nur wenige Genexpressionsprofile untersucht worden. Eine solche Patientengruppe ist die Gruppe von Patienten mit einem Tumor des

Stadiums 3 oder 4, bei dem eine bestimmte genetische Aberration, die Amplifikation des *MYCN*-Onkogens, nicht vorliegt. Für diese Patientengruppe ist die Prognoseabschätzung bislang schwierig, weshalb diese Patientengruppe besonders von einem besseren Einblick in die molekulare Biologie des Neuroblastoms oder neuen molekularen Risikomarkern profitieren würde. In dieser Arbeit wurden für diese Patientengruppe die Unterschiede in der Tumor-Genexpression in Bezug auf das Patienten-Überleben fünf Jahre nach Diagnosestellung untersucht. Zusätzlich wurde ein Verfahren der Phänotyp-Klassifikation eingesetzt, um ein prädiktives Modell zur Klassifikation der Patienten bezüglich des 5-Jahres-Überlebens zu erstellen. Um möglichst viele Genexpressionsprofile untersuchen zu können, wurden die auf unterschiedlichen Microarray-Plattformen erstellten Daten von insgesamt 66 Patienten aus zwei Studien verwendet.

Auf der Basis der Daten beider Microarray-Plattformen wurden 72 Gene mit signifikanter differentieller Expression zwischen der Gruppe der überlebenden und der verstorbenen Patienten beobachtet. Dabei wurden in der Meta-Analyse 34 Gene als signifikant differentiell exprimiert identifiziert, für welche in vergleichbaren Analysen auf den Einzeldatensätzen keine signifikant differentielle Expression gefunden wurde. Mit einem plattformübergreifenden Modell zur Prädiktion des 5-Jahres-Überlebens wurde eine Klassifikationsgenauigkeit von 77% erreicht. Der Vergleich des Expressionsverhaltens der 72 differentiell exprimierten Gene in der untersuchten Gruppe von Tumoren mit der Expression in anderen Tumoren ergab, daß die Gruppe der Tumoren des Stadiums 3 oder 4 ohne Amplifikation des *MYCN*-Gens in zwei biologische Untergruppen aufgeteilt werden kann. Eine dieser Gruppen zeigt ein Genexpressionsprofil ähnlich zu Tumoren mit einer günstigen Prognose des Stadiums 1 oder 2 ohne *MYCN*-Amplifikation, und die andere Gruppe zeigt ein Genexpressionsprofil ähnlich zu Tumoren mit einem ungünstigem Verlauf des Stadiums 3 oder 4 mit *MYCN*-Amplifikation.

Die plattform- und studienübergreifende Analyse von Genexpressionsdaten ermöglicht die Detektion von konsistenten Effekten in unabhängigen Microarray-Datensätzen. Die Meta-Analyse von Genexpressionsprofilen verschiedener Studien ermöglicht die Detektion von Genexpressionsänderungen mit erhöhter Sensitivität. Prädiktive Modelle können auf der Basis von Daten verschiedener Studien erstellt werden und werden daher auf einer größeren Anzahl an Proben generiert und validiert, was zu einer besseren Generalisierungsfähigkeit der Modelle führt.