

B Zusammenfassung

Bei manchen Krebserkrankungen ist es mit herkömmlicher Diagnostik nicht möglich, Krebs in klinisch relevante Untertypen einzuteilen. Oft unterscheiden sich aber morphologisch identisch aussehende Tumorproben stark in ihrem Ansprechen auf Medikamente, Metastasierungspotential oder Langzeitüberleben. Wäre es möglich, zuverlässig vorauszusagen, ob eine Therapie anspricht, so könnte man die vorhandenen Medikamente sinnvoller einsetzen und den Patienten eine belastende Chemotherapie ersparen.

Seit einigen Jahren ist mit der Entwicklung sogenannter Microarrays ein entscheidender Schritt in Richtung molekularer Diagnose gemacht worden. Microarrays messen die Expression Tausender Gene gleichzeitig und ermöglichen einen Einblick in die genetischen Vorgänge in den verschiedenen Geweben. Beim Vergleich von Gewebeproben ist dabei das Hauptinteresse die Identifizierung und Charakterisierung von Genen, die sich in den Gewebeproben unterschiedlich verhalten. Mithilfe solcher differentieller Gene lassen sich diagnostische Klassifikatoren auf molekularer Ebene konstruieren, die genauer als herkömmliche immunohistochemische oder histopathologische Verfahren funktionieren.

In dieser Arbeit habe ich neue Methoden und Validierungswerkzeuge zur Entwicklung von kleinen, kostengünstigen und effizienten diagnostischen Biomarkerpanels anhand von Microarraydaten größerer klinischer Studien beschrieben. Ein Biomarkerpanel kann dabei aus nur wenigen Genen bestehen und mittels z.B. qRT-PCR gemessen werden. Bei mehreren Genen bietet sich ein kleines maßgeschneidertes Microarray an. Um ein diagnostisches Biomarkerpanel zu entwickeln muss zuerst gezeigt werden, dass eine Diagnose mittels Microarrays überhaupt möglich ist. Ist das der Fall, dann wählt man sich eine geeignete Methode zur Auswahl von diskriminierenden Genen aus. Es muß allerdings noch geklärt werden, wie viele Gene und Patienten nötig sind, um eine Abschätzung der Genauigkeit des Diagnosepanels zu ermöglichen. Um schließlich die Biomarkerpanels untereinander vergleichbar zu machen, müssen diese normalisiert werden.

Im zentralen Kapitel 2 habe ich eine neue Strategie zum Design diagnostischer Microarrays eingeführt. Dabei werden in einer ersten Phase krankheitsrelevante Gene identifiziert. Einige wenige Patientenproben werden hierzu mit Microarrays, die die Expression aller Gene des menschlichen Genoms messen, analysiert. Von diesen über 20000 Genen werden die für die Diagnose relevanten Gene identifiziert und für das Design eines kleinen Biomarkerpanels ausgewählt. Mit diesem kostengünstigeren Biomarkerpanel werden in einer zweiten Phase weitere Patientenproben gemessen und ein Klassifikator trainiert. Um den Klassifizierungsverlust abzuschätzen, habe ich untersucht wie viele Patienten in der ersten Phase

nötig sind, um ein diagnostisches Microarray mit einer vorgegebenen Anzahl von Gensonden zu erstellen. Die Analyse mehrerer Genexpressionsstudien hat gezeigt, dass es schon früh möglich ist von einem großen, teuren Microarray auf ein kleines, kostengünstiges Biomarkerpanel zu wechseln. Dabei spielt die optimale Auswahl der zu untersuchenden Gene weniger eine Rolle als das anschließende Anpassen der einzelnen Gengewichte im Klassifikator. Unsere Evaluationsstrategie wendet dabei vollständige Kreuzvalidierungsmethoden an, um möglichst unverfälschte Aussagen über die Güte des Biomarkerpanels zu erlauben.

Im Kapitel 3 habe ich eine effiziente, neue Methode zur Verbesserung der Auswahl von differentiellen Genen für die Klassifizierung vorgestellt. Durch die Selektion von hoch relevanten Genen, die aber gleichzeitig untereinander nur schwach korreliert sind, konnte eine deutliche Steigerung der Klassifikationsgenauigkeit erreicht werden. Diese Verbesserung ist besonders deutlich ausgeprägt, wenn nur wenige Gene verwendet werden, wie es bei diagnostischen Biomarkerpanels zutrifft.

Im Kapitel 4 zeige ich auf, dass es mit Standardnormalisierungsverfahren für Microarrays nicht möglich ist Biomarkerpanels zu normalisieren, da diese andere Charakteristika aufweisen. Eine Normalisierung ist jedoch unabdingbar um die Ergebnisse untereinander vergleichbar machen zu können. Im Rahmen dieser Arbeit stelle ich daher zwei neue Normalisierungsverfahren für Diagnostikchips vor. Die erste Methode sucht in den Expressionsdaten nach Genen, die sich zur Normalisierung eignen. Die zweite Methode benutzt balancierte Gene und kommt ohne zusätzliche Normalisierungsgene aus. Somit können alle Gene zur Klassifikation beitragen und erlauben ein besonders effizientes Paneldesign.

Diese Erkenntnisse können in der Zukunft für die Entwicklung von kleinen Biomarkerpanels eingesetzt werden und es wird im Anhang exemplarisch anhand von zwei klinischen Anwendungen der mögliche Einsatz eines diagnostischen Panels für Genexpressionsstudien aufgezeigt.