

1 Genomics:

Identifikation neuer und bekannter Gene

OLIVER DISTLER

Unterstützt durch einen Forschungskredit der Universität Zürich.

Inhaltsverzeichnis

1.1	Einleitung	1	1.3.2	Probenherstellung und Markierung	9
1.2	Suppressive subtraktive Hybridisierung	2	1.3.3	Auswertung und Datenanalyse	10
1.2.1	RNA-Isolation und reverse Transkription	2	1.3.3.1	Messung und Erfassung der Rohdaten	10
1.2.2	Prinzip der suppressiven subtraktiven Hybridisierung	3	1.3.3.2	Normalisierung	10
1.2.3	Verifizierung differenziell exprimierter Gene	5	1.3.3.3	Reduktion falsch-positiver Signale	11
1.2.4	Limitierungen der suppressiven subtraktiven Hybridisierung	6	1.3.3.4	Statistische Analyse	12
1.2.5	Anwendungsbeispiel in der Rheumatologie	6	1.3.3.5	Speicherung der Ergebnisse in Datenbanken	14
1.3	Mikroarrays	7	1.3.4	Anwendungsbeispiel in der Rheumatologie	15
1.3.1	Herstellung des Arrays	8	1.4	Resümee	16
			1.5	Literatur	16

1.1 Einleitung

Mit der Vollendung des humanen Genomprojekts stehen Sequenzinformationen für etwa 35 000 menschliche Gene zur Verfügung (McPherson et al. 2001; Venter et al. 2001). Parallel zum menschlichen Genom konnte bis heute für zahlreiche weitere Organismen das Genom ganz oder teilweise sequenziert werden. Allerdings sind von der Mehrzahl der menschlichen Gene nur partielle Sequenzen (expressed sequence tags, EST) bekannt, wobei die meisten EST für Gene mit bisher nicht näher untersuchten biologischen Eigenschaften kodieren.

Eine Hauptaufgabe der nächsten Jahre wird daher die funktionelle Charakterisierung der sequenzierten Gene sein. In der internationalen Literatur wird hierfür oft der Begriff „functional genomics“ verwendet. Während dieser Begriff in der Literatur häufig benutzt wird, ist seine Bedeutung weit weniger klar definiert. Am ehesten wird unter „functional genomics“ die Anwendung von experimentellen molekularbiologischen Screeningverfahren zur systematischen Untersuchung einer größeren Anzahl von Genen verstanden (Hieter u. Boguski 1997). Durch die Detektion und Quantifizierung der Expression von Genen unter bestimmten experimentellen Bedingungen oder in bestimmten

Erkrankungen kann auf ihre mögliche biologische Funktion rückgeschlossen werden. Die daraus postulierten Funktionen spezifischer Gene müssen anhand adäquater Modelle verifiziert werden.

Eine klassische Herangehensweise ist in Abb. 1.1 dargestellt. Als molekularbiologische Screeningmethoden stehen neu entwickelte Verfahren wie

- Mikroarrays,
- subtraktive Hybridisierung,
- SAGE (serial analysis of gene expression) (Velculescu et al. 1995) oder
- RNA arbitrarily primed polymerase chain reaction (RAP-PCR, differential display) (Welsh et al. 1992)

zur Verfügung. Mit diesen Screeningverfahren können Gene, die in den untersuchten Bedingungen differenziell exprimiert sind, identifiziert werden. Da die Screeningverfahren eine größere Anzahl falsch-positiver Resultate produzieren können, muss die differenzielle Expression mit unabhängigen Methoden wie Real-time-PCR oder Northern-Blot verifiziert werden. „Functional genomics“ schließt auch die funktionelle Charakterisierung der identifizierten Gene ein. In Abhängigkeit von der Fragestellung kann ein Gen hierfür beispielsweise durch Transfektion mit einem Expressionsvektor überexprimiert oder durch Antisense- oder

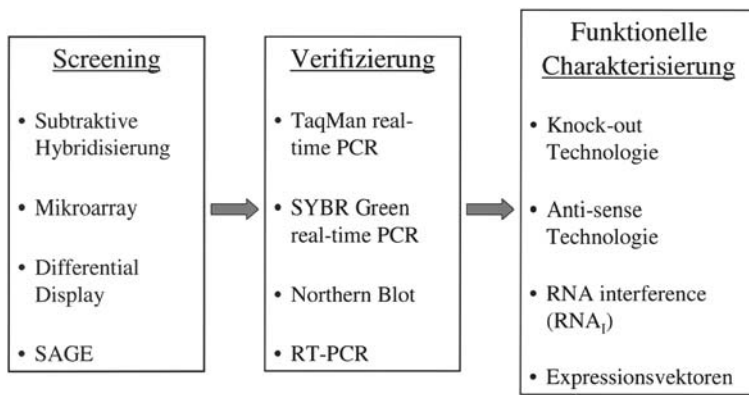


Abb. 1.1. Klassische Vorgehensweise bei molekulargenetischen Screeningverfahren. Im Anschluss an die Durchführung der Screeningverfahren zur Identifikation differenziell exprimierter Gene müssen die Ergebnisse mit unabhängigen Verfahren verifiziert werden. Der Begriff „functional genomics“ schließt auch die funktionelle Charakterisierung der identifizierten Gene in geeigneten Modellen ein. SAGE serial analysis of gene expression

Knockout-Methoden inhibiert und anschließend in einem adäquaten Modell auf seine biologischen Effekte untersucht werden.

Die molekulargenetischen Screeningverfahren sind neu entwickelte Methoden, die erst seit kurzer Zeit für eine größere Anzahl molekulargenetischer Labors verfügbar sind. Einerseits herrscht weitgehende Übereinstimmung, dass diese neuen Verfahren die molekulargenetische Forschung rasant beschleunigen und zur Identifikation wichtiger Pathomechanismen beitragen können. Andererseits besteht durch die hoch entwickelte Technologie, die spezialisiertes Fachwissen in verschiedenen Bereichen wie Molekularbiologie, Biophysik, Medizin und Bioinformatik erfordert, die Gefahr der falschen Anwendung sowie der Fehlinterpretation von Daten. Es ist nicht das Ziel dieser einleitenden Zusammenfassung, die vielfältigen Methoden für molekulargenetische Untersuchungen aufzuzählen und zu diskutieren, da dies den Rahmen dieses Kapitels schnell sprengen würde, sondern sich auf neue Strategien zu fokussieren, die wichtige Fragen in der Pathogenese rheumatischer Erkrankungen beantworten könnten. Das vorliegende Kapitel fasst daher neue Erkenntnisse über 2 häufig verwendete Verfahren zur Identifikation neuer und bekannter Gene,

- die suppressive subtraktive Hybridisierung und
- Mikroarrays,

für die rheumatologische Grundlagenforschung zusammen und zeigt Limitierungen ihrer Anwendungsmöglichkeiten auf.

1.2 Suppressive subtraktive Hybridisierung

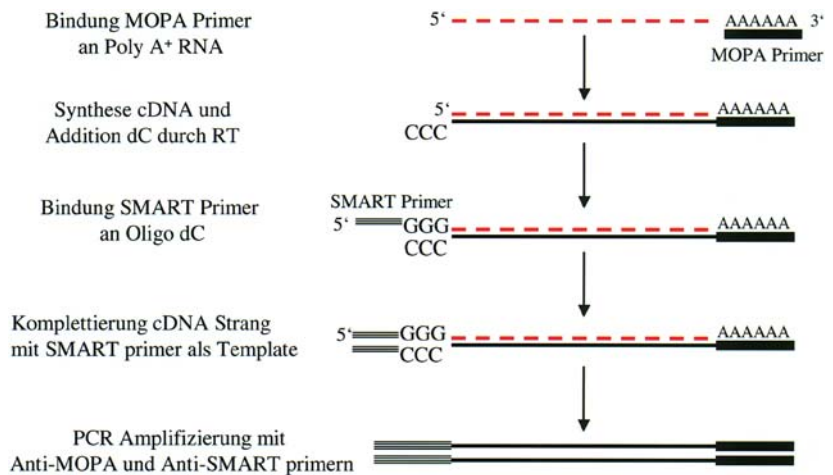
Die suppressive subtraktive Hybridisierung (SSH) ist ein Screeningverfahren zur Identifikation differenziell exprimierter Gene, das auf Techniken der Polymerasekettenreaktion (PCR) basiert (Diatchenko et al. 1996). Herkömmliche cDNA-Subtraktionsmethoden erfordern eine größere Anzahl von Hybridisierungsschritten und selektionieren nicht für Moleküle, die in geringerer Zahl transkribiert werden. Bei der SSH kann mit nur 2 Hybridisierungsschritten eine bis zu 1000fache Anreicherung unterschiedlich exprimierter Gene erreicht werden. Um auch die Identifikation seltenerer mRNA-Transkripte zu ermöglichen, wurde ein zusätzlicher Schritt in den Verlauf der Subtraktion eingebaut, in dem unter Verwendung von Standardhybridisierungskinetiken für die Expressionshäufigkeit normalisiert wird. Mit Hilfe dieser Anreicherungsmethode können auch differenziell exprimierte Gene mit niedriger Transkriptionsrate identifiziert werden.

1.2.1 RNA-Isolation und reverse Transkription

Zunächst wird Poly-A⁺-RNA aus dem Gewebe oder aus kultivierten Zellen unter Verwendung von Standardverfahren isoliert. Wird Gesamt-RNA für die Synthese von cDNA herangezogen, kommt es neben einer Umschreibung von Poly-A⁺-RNA auch zu einer Umschreibung von ribosomaler RNA. Diese Umschreibung von ribosomaler RNA ist auch durch Oligo-dT-Primer nicht völlig zu verhindern und führt zu einer ineffizienten subtraktiven Hybridisierung.

Soll dennoch Gesamt-RNA für die Experimente benutzt werden, muss für die reverse Transkription ein Verfahren verwendet werden, bei dem es zu

Abb. 1.2. SMART-Technologie zur reversen Transkription und Amplifizierung von Poly-A⁺-RNA. Die Anreicherung und reverse Transkription von Poly-A⁺-RNA erfolgt unter Verwendung von modifizierten Anti-Poly-A⁺-Primern und SMART-Primern. Abschließend erfolgt eine Amplifizierung der cDNA mittels PCR. MOPA modifizierte Anti-Poly-A⁺-Primer



einer Anreicherung von cDNA-Molekülen kommt, die von Poly-A⁺-RNA transkribiert wurden. Als Beispiel hierfür bietet sich die SMART-Technologie an (Abb. 1.2). Hierbei werden für die reverse Transkription modifizierte Anti-Poly-A⁺-Primer (MOPA) verwendet. Erreicht die reverse Transkriptase das 5'-Ende der RNA, werden durch die terminale Transferaseaktivität des Enzyms einige Desoxycytidinnukleotide am 5'-Ende angefügt. Diese Nukleotide dienen wiederum als Hybridisierungstemplate für ein spezifisches SMART-Oligonukleotid, das eine Oligo-G-Sequenz an seinem 5'-Ende besitzt. Die reverse Transkriptase benutzt nun dieses SMART-Oligonukleotid als Ablesestrang und fügt dessen Antisense-Sequenz dem abgelesenen cDNA-Einzelstrang hinzu (Zhu et al. 2001). Die spezifische Anreicherung von cDNA, die von Poly-A⁺-RNA revers transkribiert wurde, erfolgt durch PCR-Amplifizierung mit Primern gegen den initialen MOPA-Primer sowie das SMART-Oligonukleotid. Durch die Lokalisation des SMART-Oligonukleotids am 5'-Ende wird gewährleistet, dass die RNA in ihrer gesamten Länge in cDNA umgeschrieben wird.

Die Qualität und Integrität der isolierten RNA sind entscheidend für das Gelingen der Experimente. Insbesondere bei manchen Geweben wie Haut, Muskel und Sehnen kann die Isolation qualitativ hochwertiger RNA eine Herausforderung darstellen. Ein Aliquot der isolierten RNA sollte daher auf einem denaturierenden Agarose-Ethidiumbromid-Gel kontrolliert werden. Gesamt-RNA von Menschen und Säugetieren stellt sich als 2 helle Banden bei etwa 4,5 und 1,9 kb (28 S- und 18 S-RNA) dar, wobei die Intensität der Banden etwa im Verhältnis 1,5–2,5:1 liegen sollte. Teilweise oder vollständig degradierte RNA, erkenntlich an unscharfen Banden, sollte nicht für die SSH verwendet werden.

1.2.2 Prinzip der suppressiven subtraktiven Hybridisierung

Die cDNA-Population, in der differenziell exprimierte Sequenzen detektiert werden sollen, wird als Testprobe, die Referenz-cDNA-Population als Vergleichsprobe bezeichnet. Das grundlegende Prinzip der SSH besteht in einer Hybridisierung von Sequenzen, welche sowohl in der Testprobe als auch in der Vergleichsprobe exprimiert werden. Die hybridisierten Sequenzen werden anschließend entfernt. Die verbleibenden nichthybridisierten cDNA-Moleküle repräsentieren Gene, die ausschließlich in der Testprobe, aber nicht in der Vergleichsprobe vorhanden sind.

Die molekularen Details des Verfahrens sind schematisch in Abb. 1.3 dargestellt (Desaj et al. 2000). Zunächst wird die Testprobe in 2 Portionen geteilt und mit 2 unterschiedlichen Adapttern (Adapter 1 und 2) ligiert. Die Enden der Adapter sind nicht phosphoryliert, sodass nur ein Strang jedes Adapters kovalent an die 5'-Enden der cDNA-Moleküle gebunden ist. In der ersten Hybridisierung wird zu jeder Portion der Testprobe ein Überschuss an Vergleichsprobe gegeben.

Die eigentliche Hybridisierung erfolgt nach Hitze-denaturierung bei 68 °C für 6–12 h. Hierbei kommt es in jeder der 2 Portionen zur Formation verschiedener Moleküle, die in Abb. 1.3a dargestellt sind. Moleküle vom Typ A sind einzelsträngige cDNA-Moleküle der Testprobe. Aufgrund von Hybridisierungskinetiken der 2. Ordnung verlaufen Hybridisierungen von Molekülen höherer Konzentration schneller als Hybridisierungen von Molekülen mit niedriger Konzentration. Dadurch kommt es bei den Typ-A-Molekülen zu einer Anreicherung niedrig exprimierter Gene, die nun

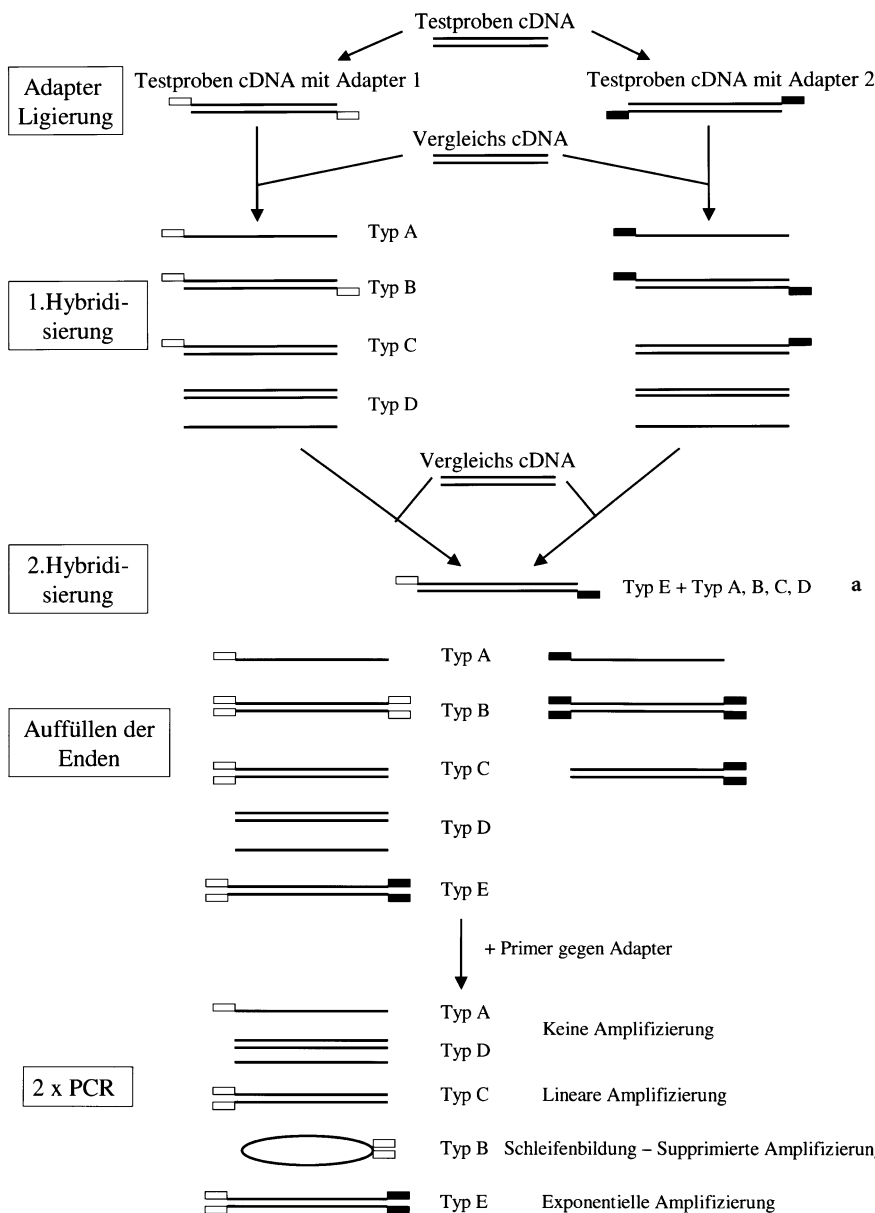


Abb. 1.3 a,b. Prinzip der suppressiven subtraktiven Hybridisierung (SSH). Das Prinzip der SSH beruht auf 2 Hybridisierungsschritten zwischen der Testprobe und der Ver-

gleichsprobe (a). Mit Hilfe einer Nested-PCR und Primern gegen initial ligierte Adapter kommt es zur exponentiellen Amplifizierung von differenziell exprimierten Sequenzen (b)

in vergleichbaren Konzentrationen zu den hoch exprimierten Genen vorliegen. Parallel findet bei den Typ-A-Molekülen eine Anreicherung von Sequenzen statt, die nur in der Testprobe vorhanden sind, da nichtdifferenziell exprimierte Sequenzen Typ-C-Moleküle mit der Vergleichsprobe bilden.

Während der 2. Hybridisierung werden die 2 Portionen der ersten Hybridisierung gemischt. Die Zugabe von frisch denaturierter Vergleichsprobe ermöglicht eine weitere Anreicherung von diffe-

renziell exprimierten Typ-A-Molekülen. Typ-A-Moleküle der beiden Portionen können nun miteinander assoziieren und Typ-B-, Typ-C- und neu Typ-E-Hybride bilden. Diese Typ-E-Hybride sind doppelsträngige Moleküle der Testprobe mit unterschiedlichen einzelsträngigen Enden, die den Adaptern 1 und 2 entsprechen.

Vor den anschließenden PCR-Amplifizierungen werden die einzelsträngigen Adapterenden aufgefüllt, sodass nun komplementäre Bindungsstellen

für spezifische PCR-Primer vorliegen (Abb. 1.3 b). Moleküle vom Typ A und D besitzen keine Bindungsstellen für die PCR-Primer und können nicht amplifiziert werden. Typ-B-Moleküle mit gleichen Adaptern an beiden Enden bilden Schleifenstrukturen, wodurch die Bindung der Primer und somit die nachfolgende Amplifizierung supprimiert (*suppressive* subtraktive Hybridisierung) werden (Lukyanov et al. 1995). Moleküle vom Typ C weisen nur eine Primerbindungsstelle auf und können deshalb nur linear amplifiziert werden. Typ-E-Moleküle besitzen dagegen 2 unterschiedliche Bindungsstellen für Primer an den jeweiligen Adapterenden, wodurch eine exponentielle Vermehrung gewährleistet ist. Als Folge werden differenziell exprimierte Moleküle in der entstehenden subtraktiven cDNA-Library stark angereichert. Die cDNA-Library kann unter Verwendung herkömmlicher Vektoren kloniert und eine beliebige Anzahl von Klonen sequenziert werden.

1.2.3 Verifizierung differenziell exprimierter Gene

Ebenso wie bei anderen Screeningverfahren entstehen bei der SSH in Abhängigkeit vom individuellen Experiment eine unterschiedlich große Anzahl falsch-positiver Resultate. In der Literatur werden hierfür Zahlen zwischen 5% und 95% angegeben, die sich mit den eigenen Erfahrungen decken (Desaj et al. 2000; Distler et al. 2002; Zhang et al. 2001). Die Rate an falsch-positiven Ergebnissen ist umso höher, je geringer die Anzahl von differenziell exprimierten mRNA-Sequenzen und je geringer die quantitativen Unterschiede in der Expressi-

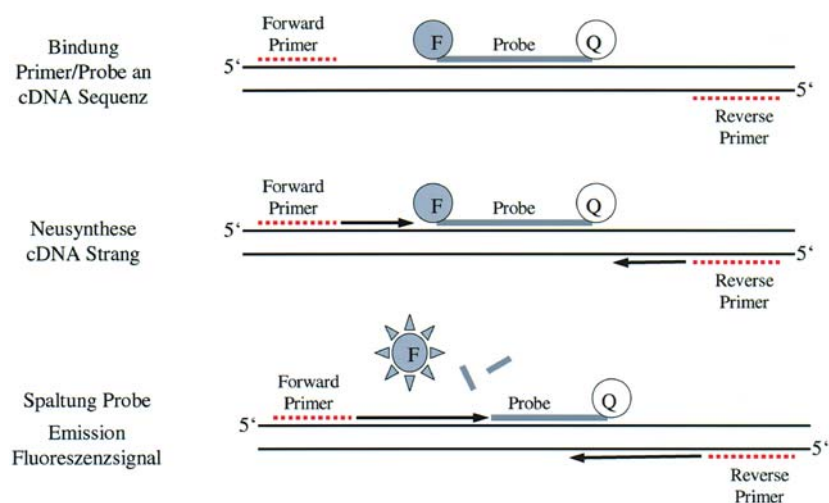
on zwischen Testprobe und Vergleichsprobe sind (Desaj et al. 2000). Eine Bestätigung der differentiellen Expression mit einem unabhängigen Verfahren ist daher unerlässlich. Neben herkömmlichen Methoden wie Northern-Blot oder RT-PCR bietet sich als Goldstandard die quantitative Real-time-PCR an.

Grundsätzlich stehen hierfür 2 Real-time-PCR-Techniken zur Verfügung:

- die TaqMan-real-time-PCR und
- die SYBR-Green-real-time-PCR.

Bei der TaqMan-real-time-PCR wird zusätzlich zu den PCR-Primern eine TaqMan-Probe verwendet, die für das zu untersuchende Molekül spezifisch ist (Abb. 1.4). Die TaqMan-Probe ist ein Oligonukleotid mit einem Fluoreszenzfarbstoff (z.B. FAM) an ihrem 5'-Ende und einem Quencher (z.B. TAMRA) an ihrem 3'-Ende (Heid et al. 1996). Bei intakter Probe findet keine Emission des Fluoreszenzfarbstoffs statt, da das Fluoreszenzsignal vom Quencher unterdrückt wird. Während der PCR bindet die TaqMan-Probe an die Zielsequenz, die von den beiden Primern flankiert wird. Da die Taq-Polymerase, die für die Real-time-PCR verwendet wird, eine 5'→3'-Exonukleaseaktivität besitzt, wird die TaqMan-Probe im Rahmen der Synthese des Neustrangs gespalten (Holland et al. 1991). Bei der Spaltung der TaqMan-Probe werden Fluoreszenzfarbstoff und Quencher getrennt, wodurch der Quencher das Fluoreszenzsignal nicht mehr unterdrücken kann. Da je synthetisiertem Neustrang ein Fluoreszenzsignal emittiert wird, ist die Höhe der Emission direkt proportional zur Anzahl der neu gebildeten Moleküle. In der exponentiellen Phase der PCR ist die Anzahl der in ei-

Abb. 1.4. Prinzip der TaqMan-real-time-PCR. Die TaqMan-real-time-PCR ist das derzeit sensitivste und spezifischste Verfahren zur Quantifizierung von mRNA. Das Prinzip beruht auf der Verwendung von sequenzspezifischen Proben, die mit einem Fluoreszenzfarbstoff markiert sind. Bei der Amplifizierung der Sequenz werden der Farbstoff abgespalten und ein Fluoreszenzsignal emittiert, F Fluoreszenzfarbstoff, Q Quencher



nem bestimmten Zyklus gebildeten cDNA-Moleküle wiederum von der Anzahl der Moleküle in der Ausgangslösung abhängig, wodurch eine quantitative Aussage über die mRNA-Expression dieses Moleküls ermöglicht wird.

Bei der SYBR-Green-real-time-PCR wird im Gegensatz zur TaqMan-real-time-PCR keine spezifische Probe verwendet (Woo et al. 1998). Die Spezifität ist daher geringer als bei der TaqMan-real-time-PCR. Allerdings ist das SYBR-Green-Verfahren durch die Einsparung einer markierten Probe wesentlich kostengünstiger und eignet sich daher insbesondere zur Bestätigung einer größeren Anzahl von Molekülen, wie dies bei der SSH der Fall ist. Das Prinzip dieses Verfahrens beruht auf der Eigenschaft von SYBR Green, explizit an doppelsträngige DNA zu binden und dabei ein Fluoreszenzsignal zu emittieren. Ähnlich wie bei der TaqMan-real-time-PCR ist die Intensität des Signals proportional zur Anzahl der neu gebildeten (doppelsträngigen) Moleküle und somit zur Menge des Moleküls in der Ausgangslösung. Da SYBR Green an alle doppelsträngigen DNA-Moleküle bindet, kann es bei dieser Methode zu unspezifischen Signalen aufgrund von Primerdimeren oder unspezifischen PCR-Produkten kommen (Vandesompele et al. 2002). Im Rahmen der Auswertung muss daher für diese Möglichkeiten mit einem Ethidiumbromid-Agarose-Gel (unspezifische PCR-Produkte) und einer Dissoziationsanalyse (Primerdimere dissoziieren bei einer anderen Temperatur als das spezifische PCR-Produkt) kontrolliert werden.

1.2.4 Limitierungen der suppressiven subtraktiven Hybridisierung

Bei der subtraktiven Hybridisierung besteht die Möglichkeit, unbekannte Gene bzw. Gene mit noch unbekannter Funktion zu identifizieren. Erfahrungsgemäß werden in größeren Ansätzen etwa 10–30% differenziell exprimierte Gene mit noch unbekannter Funktion detektiert (Desaj et al. 2000; Distler et al. 2002; Zhang et al. 2001). Auf der anderen Seite ist das Erstellen eines kompletten Expressionsprofils mit Hilfe der SSH sehr kosten- und zeitintensiv. Untersuchungen zeigen, dass in Abhängigkeit vom Versuchsansatz 300–500 Klone analysiert werden müssen, um alle differenziell exprimierten Moleküle – auch solche mit niedriger Expressionsrate – zu erfassen (Mueller et al. 1995; von Stein et al. 1997). Viele Klone werden dabei mehrfach gepickt, wodurch sich eine gewisse Redundanz der SSH zeigt.

Weiterhin muss bedacht werden, dass für die SSH eine Test- und Vergleichsprobe von nur einem Patienten bzw. einer gesunden Vergleichsperson herangezogen wird. Um zu allgemeingültigen Aussagen zu gelangen, muss die differenzielle Expression der identifizierten Gene an weiteren Patienten bzw. Proben bestätigt werden. Wird für die SSH Gewebe verwendet, das aus sehr heterogenen Zellpopulationen besteht, nimmt die Effizienz der SSH deutlich ab (Desaj et al. 2000). Homogenes Gewebe mit einer großen Anzahl gleichartiger Zielzellen, wie z.B. einige Tumorarten oder aber kultivierte Zellen, eignen sich dagegen besser als Ausgangsmaterial für die SSH. Dies gilt in vergleichbarem Maß auch für Mikroarrays und andere Screeningverfahren.

1.2.5 Anwendungsbeispiel in der Rheumatologie

Infektiöse Erreger werden seit längerem als mögliche Faktoren in der Pathogenese der rheumatoiden Arthritis (RA) diskutiert. In diesem Zusammenhang sind endogene retrovirale Sequenzen von besonderem Interesse. Kürzlich konnte gezeigt werden, dass das human LINE-1(L1)-Element, bestehend aus einem ORF 1/p40 (ORF: open reading frame) und einem ORF 2, in synovialen Biopsien von Patienten mit RA exprimiert wird (Neidhart et al. 2000). Synoviale Biopsien von Patienten mit Osteoarthritis (OA) sowie von gesunden Kontrollen zeigten keine oder nur eine minimale Expression dieser endogenen retroviralen Sequenz. Um zu untersuchen, welche Stoffwechselwege durch die Expression von L1 induziert werden, wurden kultivierte synoviale Fibroblasten von RA-Patienten mit einem Expressionsvektor für das L1-Element transfiguriert. Differenziell exprimierte Gene im Vergleich zu nicht transfigurierten Fibroblasten wurden mit Hilfe der SSH identifiziert. Neben anderen Sequenzen zeigte sich die SAPK2 δ (stress-activated protein kinase 2 δ) nach Transfektion mit L1 induziert, während bei leervektortransfigurierten synovialen Fibroblasten keine Induktion dieses Moleküls zu beobachten war. Die differenzielle Expression konnte mittels konventioneller RT-PCR in weiteren Versuchen bestätigt werden. Die mögliche Bedeutung der SAPK2 δ für die Pathogenese der RA wurde weiter durch das Expressionsmuster untermauert. Während Synovialbiopsien von OA-Patienten keine Expression der SAPK2 δ aufwiesen, zeigte sich bei Patienten mit RA eine deutliche Expression an Orten der Gelenkdestruktion in Koexpression zu L1. Anschlussexperimente deuten zudem darauf hin,

dass über Aktivierung der SAPK2 δ in synovialen Fibroblasten eine Induktion von Matrixmetalloproteinasen (MMP) stattfindet. Humane L1-Elemente könnten daher über eine Induktion der SAPK2 δ und der Matrixmetalloproteinasen an der Gelenkdestruktion der RA beteiligt sein.

1.3 Mikroarrays

Nach Vollendung des humanen Genomprojekts besteht die Hauptaufgabe der kommenden Jahre in der Erforschung der biologischen Funktionen der identifizierten Gensequenzen. In diesem Zusammenhang stellen die Weiterentwicklung und Optimierung der Genarraytechnologie die wesentlichsten technischen Fortschritte der letzten Jahre dar. Die große Anzahl von Gensequenzen auf einem Array ermöglicht die Erstellung eines umfassenden Expressionsprofils unter verschiedenen experimentellen Bedingungen. Neben zahlreichen weiteren Anwendungsmöglichkeiten kann die Arraytechnologie in der Pathogeneseforschung auch als Screeningverfahren für die Identifikation von Schlüsselmolekülen verwendet werden. Einer der Nachteile von Mikroarrays besteht in den hohen Kosten, welche die Anwendung in der akademi-

schen Forschung einschränken. An akademischen Zentren werden daher derzeit zentrale Forschungseinrichtungen aufgebaut, welche benötigte Gerätschaften und Materialien bereitstellen und die Durchführung der Arrayexperimente sowie die Auswertung der Daten fachlich begleiten. Aufgrund der breiten Anwendungsmöglichkeiten ist die Arraytechnologie auch ein wichtiger wirtschaftlicher Faktor für die Biotechnologiebranche. Nach Schätzungen wird der Umsatz für Mikroarrays weltweit von 531 Mio. US \$ im Jahre 2000 auf etwa 3,3 Mrd. US \$ im Jahr 2004 steigen (Frost u. Sullivan 2001).

Grundsätzlich stehen 2 größere DNA-Mikroarraytechnologien zur Verfügung, bei denen entweder Oligonukleotide oder cDNA auf die Oberfläche von Glaträgern oder Nylonmembranen aufgebracht werden. In Abb. 1.5 sind die einzelnen Schritte am Beispiel eines auf cDNA-Sequenzen basierenden Glasarrays zusammengefasst (Mousses et al. 2000). Mikroarrays können kommerziell erstanden oder selbst hergestellt werden. Hierfür werden mittels PCR amplifizierte und anschließend aufgereinigte cDNA-Fragmente auf eine präparierte Glasoberfläche aufgetragen. In Abhängigkeit vom Verfahren können hierbei bis zu mehrere Zehntausend cDNA-Fragmente auf eine Glasoberfläche in der Größe eines Objektträgers platziert werden (Lockhart u. Winzeler 2000).

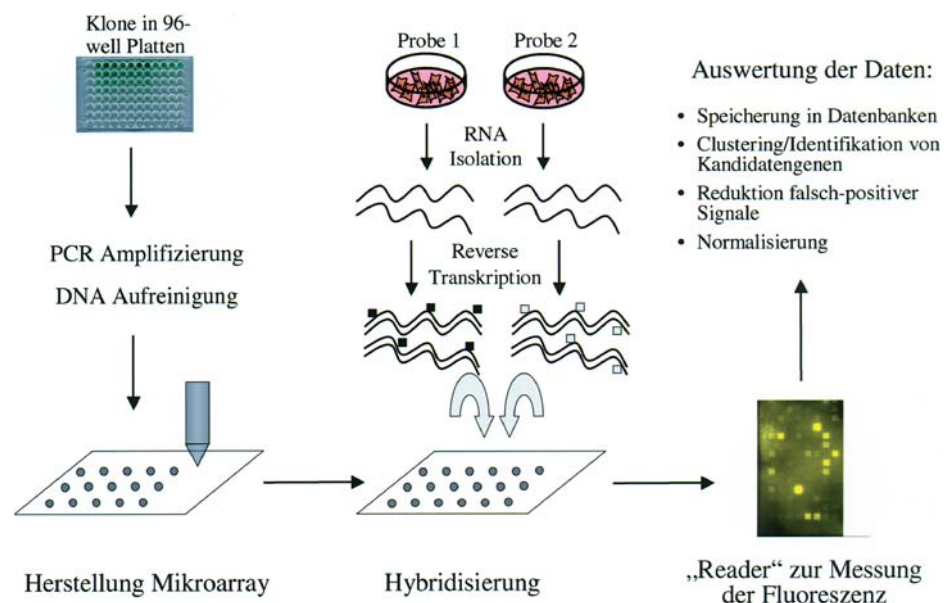


Abb. 1.5. Ablauf von Mikroarrayexperimenten. Mikroarrays können kommerziell erstanden oder selbst hergestellt werden. Nach RNA-Isolation und Markierung im Rahmen der reversen Transkription erfolgt die Hybridisierung mit der

Oberfläche des Trägermaterials. Nach stringenten Waschschritten wird das Fluoreszenzsignal abgelesen und ein Rohdatenbild erzeugt. Abschließend erfolgt die statistische Auswertung der Daten

Für jedes Experiment wird RNA aus 2 zu vergleichenden Zellpopulationen oder Geweben isoliert. Bei der anschließenden reversen Transkription werden die beiden cDNA-Proben mit 2 unterschiedlichen Fluoreszenzfarbstoffen (z. B. Cy3 und Cy5) markiert. Die beiden markierten Proben werden dann zusammen auf die Glasoberfläche aufgebracht und hybridisieren dort mit den aufgetragenen cDNA-Fragmenten. Ist eine spezifische RNA in einer der Proben häufiger vertreten, hybridisiert eine größere Menge der entsprechend markierten cDNA an die korrespondierenden cDNA-Fragmente auf der Glasoberfläche. Das Verhältnis der Fluoreszenzintensität zwischen den beiden verwendeten Fluoreszenzfarbstoffen ermöglicht dann eine Aussage über die Unterschiede in der Expressionshäufigkeit des untersuchten Moleküls. Alternativ können die Proben auch radioaktiv markiert und Filtermembranen als Trägermaterial für die Arrays verwendet werden. Die Hybridisierung erfolgt dann meist parallel auf unterschiedlichen Arrays oder hintereinander nach „Stripping“ auf derselben Membran. Aufgrund der riesigen Datenmengen bei Mikroarrays müssen zur Auswertung der Ergebnisse verschiedene Verfahren aus der Bioinformatik angewendet werden.

1.3.1 Herstellung des Arrays

An erster Stelle steht die Auswahl der Sequenzen, die auf dem Array untersucht werden sollen. Eine große Anzahl verschiedener Sequenzen können als Klone in *E.-coli*-Kulturen in 96-Loch-Platten vom IMAGE-Konsortium (Integrated Molecular Analysis of Genomes and their Expression, <http://image.llnl.gov>) erworben werden. Neben bekannten Sequenzen stehen auch EST (expressed sequence tags) zur Verfügung, die nach überlappenden Sequenzen in Cluster eingeteilt sind (Bowtell 1999). Die Plasmid-DNA kann aus den *E.-coli*-Kulturen unter Verwendung von Miniprep-Kits isoliert und als Template für eine PCR-Amplifizierung mit vektor- oder genspezifischen Primern verwendet werden. Die Herstellung der *E.-coli*-Klone verläuft automatisiert, dennoch sind bei deren Verarbeitung größtmögliche Vorsichtsmaßnahmen nötig, um eine Kontamination durch Inhalte benachbarter Löcher zu vermeiden (Knight 2001).

Die aufgereinigten PCR-Produkte werden unter Verwendung eines „Printers“ (Arrayer) auf die Trägeroberfläche des Arrays aufgebracht (Bowtell 1999; Mousses et al. 2000). Bei dem Printer handelt es sich um einen PC-gesteuerten Präzisions-

roboter, dessen Arm sich in xyz-Achse mit höchster Genauigkeit bewegen kann. Das Ansaugen der cDNA aus den 96-Loch-Platten erfolgt über Stifte aus Edelstahl, die in der Mitte einen etwa 0,002 mm großen Schlitz besitzen, in den minimale Flüssigkeitsmengen (250–500 nl) aufgenommen werden können. Die Apparatur transportiert die Stifte dann auf die Trägeroberfläche des Arrays, auf dem wenige Nanoliter der cDNA aufgebracht werden. Auf diese Weise entstehen auf der Glasoberfläche cDNA-Spots mit einem Durchmesser von 100–150 µm, der Abstand zum nächsten cDNA-Spot beträgt etwa 200–250 µm (Cheung et al. 1999). Anschließend werden die Stifte mit destilliertem Wasser gewaschen und getrocknet, ehe der Vorgang für neue cDNA-Volumina wiederholt wird. Heutige Printer bewältigen etwa 7 Spots/s. Auf diese Weise können 42 Glasarrays mit je 4000 Genen in weniger als 8 h produziert werden.

Glas hat als Trägermaterial mehrere Vorteile gegenüber Nylonmembranen. Bei der undurchlässigen und glatten Oberfläche von Glas werden die Hybridisierungskinetiken weniger beeinflusst als bei den porösen Nylonmembranen, bei denen die Proben und Waschflüssigkeiten erst in die Poren diffundieren müssen. Da die Lokalisation der cDNA-Fragmente oder Oligonukleotide aus dem gleichen Grund präziser definiert ist als auf Nylonmembranen, ist die abschließende Bildverarbeitung einfacher und besser reproduzierbar (Southern et al. 1999). Bei Fluoreszenzverfahren bestehen zudem geringere Probleme mit Hintergrundsignalen, und es können 2 unterschiedlich markierte Proben auf einem Array in einer Hybridisierung untersucht werden (Cheung et al. 1999). Dennoch haben sich neben Glas auch Nylonmembranen bei richtiger Anwendung und Interpretation als adäquate Trägersubstanz für Mikroarrays etabliert.

Eine Alternative zu cDNA-Arrays stellen Oligonukleotidarrays dar. Oligonukleotide können ebenso wie cDNA-Fragmente extern synthetisiert und auf die Glasoberfläche aufgebracht oder aber direkt auf dem Glas synthetisiert werden. Mit der 2. Methode lässt sich eine höhere Dichte auf dem Array erreichen. Um sterische Interaktionen zwischen den dicht gepackten Oligonukleotiden zu verhindern und um die Hybridisierung mit den Proben zu erleichtern, werden die Oligonukleotide über Oligoethylenglykole mit der Glasoberfläche verbunden (Southern et al. 1999). Bei längeren Oligoethylenglykolen verbessern sich die Hybridisierungseigenschaften mit der Probe, da auf diese Weise Sequenzen, die näher an der Glasoberflä-

che lokalisiert sind, besser für die Hybridisierung zugänglich gemacht werden.

Der am häufigsten verwendete Oligonukleotid-glasarray beruht auf der Genchiptechnologie von Affymetrix (Santa Clara, Kalifornien, USA). Die Oligonukleotidsynthese erfolgt in situ auf der Glasoberfläche unter Verwendung photolithographischer Techniken. Durch gezielte Belichtung eines Areals werden photolabile Schutzgruppen von der Oberfläche entfernt. Anschließend kann spezifisch an den beleuchteten Stellen die Ankoppelung eines Nukleotids erfolgen. Dieser Prozess wird mehrfach wiederholt, sodass schließlich definierte 25-mer-Oligonukleotide entstehen (Lipshutz et al. 1999). Um eine ausreichende Spezifität zu gewährleisten, sind pro Gen mehrere Oligonukleotide auf dem Array lokalisiert, die zu verschiedenen Abschnitten der Gensequenz homolog sind. Durch diese Redundanz wird die Identifikation von falsch-positiven und falsch-negativen Resultaten erleichtert und für die verschiedenen Hybridisierungskinetiken der einzelnen Oligonukleotide gemittelt. Als Negativkontrolle wird darüber hinaus pro Gensequenz ein Mismatch-Oligonukleotid verwendet, bei dem ein einzelnes Nukleotid im Vergleich zum Perfect-match-Oligonukleotid ausgetauscht ist.

1.3.2 Probenherstellung und Markierung

Die Extraktion qualitativ hochwertiger RNA ist einer der kritischsten Schritte bei Arrayexperimenten. Auch bereits geringfügig degradierte RNA-Proben, die für andere molekularbiologische Anwendungen durchaus verwendet werden können, führen bei Arrays häufig zu unvollständigen und nicht reproduzierbaren Hybridisierungssignalen. Die RNA-Proben sollten daher grundsätzlich auf ein denaturierendes RNA-Agarosegel geladen und nach ihrer Integrität beurteilt werden (s. Abschnitt 1.2 „Suppressive subtraktive Hybridisierung“). Verunreinigungen wie z.B. zelluläre Proteine, Lipide oder Karbohydrate können darüber hinaus eine unspezifische Bindung von fluoreszenzmarkierten cDNA-Fragmenten auf der Arrayoberfläche verursachen.

Ebenso wie bei anderen Screeningverfahren schränken heterogene Gewebe die Aussagekraft der Analyse stark ein. Eine differenzielle Expression eines Moleküls kann allein durch ein zufälliges Übergewicht einer bestimmten Zellpopulation in einer der beiden Proben begründet sein. Umgekehrt kann die differenzielle Expression eines Mo-

leküls, das nur in einer bestimmten Zellart im Gewebe differenziell exprimiert wird, durch die identische Expression in anderen Zellarten überdeckt werden.

Bei heterogenen Geweben stellt daher die Lasermikrodissektion eine Lösungsmöglichkeit dar. Hierbei werden immunhistochemisch oder anderweitig markierte Zellen unter Verwendung eines Präzisionslasers aus dem Gewebe herausgeschnitten. Dieser Vorgang wird an mehreren Parallelschnitten wiederholt, anschließend wird aus den herausgeschnittenen Zellen RNA extrahiert (Lechner et al. 2001). Auf diese Weise sind auch Aussagen über unterschiedliche Gewebeabschnitte, z.B. „lining layer“ und „sublining“ bei der RA, möglich. Darüber hinaus sind Zelllinien sehr gut als Modell für Mikroarrayexperimente geeignet, da sie ausreichende Mengen RNA liefern und zudem eine wesentlich homogenere Population darstellen als primäre Zellkulturen (Bowtell 1999).

Ein häufiges Problem bei den Versuchen ist die limitierte Menge an RNA, die für die Experimente zur Verfügung steht. Dies trifft insbesondere auf Studien zu, die mit menschlichen Geweben arbeiten, da hierbei auch ethische Gesichtspunkte zu berücksichtigen sind. Ein Schwerpunkt der Entwicklung der letzten Jahre lag daher auf Methoden zur Amplifizierung von niedrig konzentrierter Ausgangs-RNA.

Einige dieser Verfahren beruhen auf einer Amplifizierung mit PCR-basierten Techniken. Hierfür kann beispielsweise die oben beschriebene SMART-Technologie oder die RAP-PCR verwendet werden (Lockhart u. Winzeler 2000; Neumann et al. 2002). Eine weitere Strategie stellt die Signalamplifizierung nach erfolgter Hybridisierung unter Verwendung markierter Antikörper dar (Schulze u. Downward 2001). Bei diesen Verfahren besteht allerdings die Möglichkeit, dass sich die Zusammensetzung der RNA-Population durch ungleichmäßige Amplifizierung individueller Gene verändert. Um derartige systematische Fehler auszuschließen, müssen bei diesen Verfahren entsprechende Kontrollen mitgeführt werden. Insbesondere ist es nicht möglich, amplifizierte Proben mit nicht behandelten Proben zu vergleichen.

Für die Markierung der Proben werden am häufigsten Cy3-dUTP und Cy5-dUTP verwendet. Diese Fluoreszenzfarbstoffe haben eine gute Photostabilität und deutlich getrennte Emissionsspektren (Duggan et al. 1999). Bei der Markierung ist darauf zu achten, dass beide Farbstoffe im Rahmen der reversen Transkription mit vergleichbarer Effizienz inkorporiert werden. Hierfür stehen speziell

entwickelte reverse Transkriptasen zur Verfügung (Li et al. 2002). Bei Nylonarrays werden die Proben meist radioaktiv mit $\alpha^{32}\text{P}$ - oder $\alpha^{33}\text{P}$ -dATP markiert. Die $\alpha^{32}\text{P}$ -Isotope sind etwa 4fach sensibler als $\alpha^{33}\text{P}$ -Isotope, allerdings können bei hoch exprimierten Molekülen Überstrahlungen in benachbarte Areale auftreten. Bei radioaktiv markierten Proben erfolgt die Auswertung mittels eines Phosphorimagers nach unterschiedlichen Expositionszeiten. Bei der Verwendung von fluoreszenzmarkierten Proben wird die Auswertung mit konfokalen Lasermikroskopen durchgeführt.

1.3.3 Auswertung und Datenanalyse

Ein häufig geäußertes Argument in der Diskussion um Vor- und Nachteile von Mikroarrays sind die großen Mengen an Daten, die bei Arrayexperimenten entstehen und die von Wissenschaftlern nicht in sinnvoller Weise ausgewertet und interpretiert werden könnten. Tatsächlich erfordern die vorher in der molekularbiologischen Forschung nicht gekannten riesigen Datenvolumen neue Verfahren der statistischen Auswertung sowie der Speicherung in Datenbanken, die allgemein zugänglich sein müssen (Lockhart u. Winzler 2000). Die Entwicklung dieser Verfahren kann nur durch eine enge Zusammenarbeit zwischen verschiedenen Disziplinen wie Bioinformatik, Biostatistik, Biophysik und Molekularbiologie erreicht werden. Weitgehende Einigkeit herrscht darüber, dass im Rahmen der Mikroarraytechnologie in diesem Bereich derzeit der größte Entwicklungsbedarf liegt (Brazma u. Vilo 2000; Li et al. 2002; Wu 2001). Hierbei kann auch auf Erfahrungen in anderen Bereichen wie der Flugindustrie oder der Plasmaphysik zurückgegriffen werden, die Methoden zur Analyse ähnlich großer Datenmengen entwickelt haben (Lockhart u. Winzler 2000). Die Auswertung und Datenanalyse umfasst folgende Bereiche:

- Messung und Erfassung der Rohdaten
- Normalisierung der Expressionsdaten
- Identifikation von falsch-positiven Signalen
- statistische Analyse
- Speicherung der Ergebnisse in Datenbanken

Ein allgemein gültiges Verfahren zur Auswertung von Arraydaten existiert nicht, da die Auswertung von der Art des Arrays, von der experimentellen Fragestellung, vom experimentellen Aufbau sowie von individuellen laborspezifischen Parametern abhängig ist. Darüber hinaus sind optimale statistische Algorithmen für die jeweiligen Fragestellun-

gen erst in der Entwicklung. Die folgenden Abschnitte fassen grundlegende Richtlinien zu den jeweiligen Bereichen der Datenanalyse zusammen.

1.3.3.1 Messung und Erfassung der Rohdaten

Der erste Schritt in der Datenauswertung stellt die Lokalisation der einzelnen Signale auf dem Array dar. In Abhängigkeit von der verwendeten Software und der Art des Arrays wird ein Gitter über die Signale gelegt, das manuell konfiguriert werden kann. Hierbei sollte darauf geachtet werden, dass die Messfläche identisch mit der Signalfäche ist, da sonst die Stärke des Signals nicht vollständig erfasst wird. Vorsicht ist bei Signalen mit niedriger Intensität geboten, da hierbei häufig fälschlicherweise vielfache Erhöhungen im Vergleich zur Kontrollprobe detektiert werden, ohne dass diese Ergebnisse reproduzierbar sind. Oft werden daher Signale, deren Intensität geringer ist als der Mittelwert plus 2-mal die Standardabweichung der Hintergrundsignale, von der weiteren Analyse ausgeschlossen (Li et al. 2002). Eine weitere wichtige Aussage über die Homogenität und Qualität eines Signals liefert die Intensität individueller Pixel innerhalb des Signalbereichs, aus denen die Standardabweichung eines Signals berechnet werden kann. Gleiches gilt für die Berechnung der Hintergrundsignale (Brazma u. Vilo 2000).

1.3.3.2 Normalisierung

Bevor Aussagen über die differenzielle Expression bestimmter Moleküle getroffen werden können, müssen die Werte in den Proben für unterschiedliche Mengen an Ausgangs-RNA abgeglichen werden (Li et al. 2002; Schulze u. Downward 2001; Wu 2001).

Bei der globalen Normalisierung wird die absolute Signalintensität aller auf dem Array befindlichen Gene für die Berechnung der geladenen RNA-Menge herangezogen. Da zwischen den 2 Vergleichsproben immer nur ein kleiner Prozentsatz von Molekülen differenziell exprimiert wird, fallen diese bei der Berechnung der absoluten Intensität nicht ins Gewicht. Dieses Verfahren ist insbesondere für Arrays mit einer großen Anzahl verschiedener Gene geeignet sowie für Experimente mit geringen Unterschieden in den Vergleichsproben.

Bei Arrays mit einer geringeren Anzahl von Genen wie z.B. Nylonmembranarrays werden meist interne Kontrollen (housekeeping genes) zur Normalisierung herangezogen. Da bei diesem Verfah-

ren nur wenige Signale ausgewertet und zur Berechnung aller weiteren Werte herangezogen werden, ist hier besonders auf die Signalqualität (Standardabweichung der Signalintensität) zu achten. Darüber hinaus muss bedacht werden, dass sich auch die Expression der so genannten „house-keeping genes“ unter bestimmten experimentellen Bedingungen verändern kann. Ein Beispiel hierfür ist die Beeinflussung von G3PDH durch Hypoxieexposition.

Neben unterschiedlich geladenen Mengen an RNA können auch Faktoren eine Rolle spielen, welche die Intensität der Signale nichtlinear beeinflussen. Beispielsweise kann es im Rahmen der Hybridisierung zu Sättigungsprozessen kommen, bei denen hoch exprimierte mRNA-Moleküle nicht durch eine proportional hohe Signalintensität repräsentiert werden, da nur eine begrenzte Zahl von cDNA-Molekülen dieser Gene auf dem Array für die Hybridisierung zur Verfügung steht. Statistische Verfahren zur Berücksichtigung dieser nichtlinearen Effekte befinden sich derzeit in der Entwicklung (Wu 2001).

1.3.3.3 Reduktion falsch-positiver Signale

Die Identifikation falsch-positiver Signale stellt ein besonderes Problem bei der Auswertung der Mikroarrayexperimente dar. Beim Vergleich von 2 unterschiedlichen mRNA-Populationen mittels Oligonukleotidarray beträgt die Rate an falsch-positiven Ergebnissen etwa 1–2% (Lipshutz et al. 1999; Mills et al. 2001). Auch wenn dieser Prozentsatz zunächst recht niedrig erscheint, muss bedacht werden, dass dies beispielsweise bei einem Array mit 10 000 Genen eine Anzahl von 100 falsch-positiven Ergebnissen zur Folge hat. Da bei vielen Experimenten oft nur einige hundert Gene verändert gefunden werden, kann die Anzahl falsch-positiver Ergebnisse die Auswertung der Arraydaten stark erschweren.

Die Wahrscheinlichkeit falsch-positiver Ergebnisse korreliert mit der Höhe der differentiellen Expression. Bei Molekülen mit einer bis zu 2fachen Veränderung zwischen den mRNA-Populationen sind falsch-positive Ergebnisse deutlich wahrscheinlicher als bei Molekülen mit größeren Expressionsunterschieden. In zahlreichen Studien wurde daher die Schwellenwertmethode zur Reduktion falsch-positiver Ergebnisse eingesetzt, indem nur eine differenzielle Expression mit einer mehr als 2fachen Veränderung in die weitere Analyse einbezogen wurde (Wu 2001).

Allerdings ist dieser beliebig festgelegte Schwellenwert mit statistischen Modellen nicht zu be-

gründen, da die Reproduzierbarkeit der Expressionsunterschiede von zahlreichen labor- und instrumentenspezifischen Parametern beeinflusst wird und keinem starren Schema folgt. Während daher in einigen Laboratorien eine 2fache Veränderung einen optimalen Schwellenwert darstellen kann, kann dieser Wert in anderen Laboratorien völlig unzureichend sein (Li et al. 2002). Ein weiteres Problem eines beliebig festgelegten Schwellenwerts besteht darin, dass bei diesem Verfahren eine Anzahl richtig-positiver Moleküle von der weiteren Analyse ausgeschlossen wird, die trotz einer weniger als 2fachen Erhöhung wichtige biologische Funktionen erfüllen können. Besondere Vorsicht ist daher bei Ergebnissen für Moleküle geboten, deren cDNA zwar auf dem Array vorhanden ist, für die aber mit dem Schwellenwertverfahren keine differenzielle Expression detektiert werden konnte. Auf der anderen Seite des Spektrums findet sich auch ein signifikanter Prozentsatz falsch-positiver Ergebnisse bei Molekülen mit einer mehr als 2fachen Erhöhung (Churchill u. Oliver 2001).

Eine alternative Vorgehensweise zur Reduktion falsch-positiver Ergebnisse ist die Wiederholungsanalyse. Hierbei kann zwischen Wiederholungen der Arrayhybridisierung (analytische Wiederholung) und Wiederholung des Experiments (biologische Wiederholung) unterschieden werden (Mills et al. 2001). Bei der analytischen Wiederholung werden RNA-Proben aus dem gleichen Experiment mit verschiedenen Nylonmembranen oder Glaträgern hybridisiert und nur solche Gene als verändert gewertet, die auch in den wiederholten Versuchsansätzen eine differenzielle Expression auf den verschiedenen Arrays zeigen. Bei der biologischen Wiederholung wird das biologische Experiment (z. B. Stimulierung von Zellen mit einem Wachstumsfaktor) wiederholt und die extrahierte RNA bzw. cDNA jeweils mit Arrays hybridisiert. Entsprechend der analytischen Wiederholung werden nur solche Gene als verändert gewertet, die reproduzierbar eine unterschiedliche Expression in den Experimenten zeigten. Aufgrund der mittlerweile gut etablierten Arraytechniken ist die biologische Variabilität häufig deutlich größer als die analytische Variabilität. Daher ist bereits bei der Planung und Durchführung der Experimente auf eine bestmögliche Standardisierung der Versuchsbedingungen zu achten (Schulze u. Downward 2001; Wu 2001). Der Nachteil dieses Verfahrens besteht in den hohen Kosten, die infolge der wiederholten Hybridisierungen entstehen. Ein Beispiel für einen systematischen Ansatz mit dieser Herangehensweise wurde von Mills et al. (2001) geschildert.

1.3.3.4 Statistische Analyse

Nach Messung und Erfassung der Rohdaten, Normalisierung sowie Reduktion falsch-positiver Ergebnisse besteht die nächste Aufgabe darin, die erhaltenen Daten nach möglichen funktionellen Zusammenhängen zu katalogisieren und somit einer Interpretation zugänglich zu machen. Die verschiedenen Analyseverfahren können dabei grundsätzlich in 2 Gruppen eingeteilt werden:

- Gruppierung von Genen (Clustering) und
- Filterverfahren zur Identifikation von Kandidatengen (Wu 2001).

Ziel der Clusterverfahren ist die Anordnung von Genen mit gleichem Expressionsverhalten in Gruppen, um so Rückschlüsse auf ähnliche Regulationsmechanismen oder biologische Funktionen zu ziehen. Bei den Filterverfahren ist die Identifikation von typischen Effektormolekülen eines bestimmten Stimulus (z.B. die molekularen Effekte eines Wachstumsfaktors auf eine bestimmte Zellart) Ziel der Analyse.

Filterverfahren zur Identifikation von Kandidatengen

Die Wahl der Methode zur Identifikation von Kandidatengen ist weitgehend abhängig vom Aufbau des Experiments und den zur Verfügung stehenden Daten. Ein einfaches und zugleich häufig durchgeführtes Experiment ist der Vergleich von 2 Konditionen, beispielsweise von stimulierten und nichtstimulierten Zellen.

Neben den bereits angesprochenen Richtlinien zur Normalisierung der Daten und zur Reduktion falsch-positiver Ergebnisse muss auch das Problem des Tests multipler Hypothesen beachtet werden. Bei der Anwendung von statistischen Verfahren mit einem großen Datenvolumen kommt es allein aufgrund der großen Anzahl von durchgeführten Tests zu einer bestimmten Anzahl von p -Werten, welche die statistische Signifikanzgrenze erreichen. Die p -Werte einer statistischen Analyse von Mikroarrays müssen daher einer Korrektur unterzogen werden. Am häufigsten wird in diesem Zusammenhang die Bonferroni-Korrektur verwendet, bei welcher der angestrebte p -Wert durch die Anzahl der Tests (Anzahl der Gene) geteilt wird. Allerdings resultiert dieses Verfahren in zu strengen p -Werten für das individuelle Gen, sodass neuere Verfahren wie die Bootstrap-Methode angemessener erscheinen (Wu 2001).

Zur Reduktion falsch-positiver Ergebnisse sollten Experimente und auch Hybridisierungen wie-

derholt werden. Die Resultate der wiederholten Experimente können mittels parametrischer oder nicht parametrischer Test statistisch analysiert werden. Der t-Test als Prototyp eines parametrischen Tests setzt eine Normalverteilung der Daten voraus, in diesem Beispiel eine Normalverteilung der Expressionswerte eines bestimmten Gens in den wiederholten Experimenten. Eine Normalverteilung der Daten wird bei Mikroarrayexperimenten meist ohne weitere Beweisführung vorausgesetzt, da eine Normalverteilung einerseits aufgrund der meist niedrigen Anzahl an Wiederholungen schwierig nachzuweisen und andererseits aufgrund der Homogenität der wiederholten Experimente anzunehmen ist (Wu 2001).

Der t-Test ist ein häufig angewandtes Verfahren zur statistischen Analyse von Mikroarrays. Soll keine Normalverteilung der Daten vorausgesetzt werden, können die Ergebnisse auch mit nicht parametrischen Tests wie dem Mann-Whitney-Test analysiert werden. Hier basiert der Vergleich nicht wie beim t-Test auf der Varianz der Experimente, sondern auf Unterschieden in der Rangreihenfolge, die jedem Experiment zugeordnet wird.

Bisher wurden nur Experimente angesprochen, in denen 2 verschiedene Bedingungen verglichen wurden. Häufig interessiert aber auch der Vergleich mehrerer Bedingungen. Beispielsweise können Synovialbiopsien von Patienten mit rheumatoider Arthritis mit Synovialbiopsien von Patienten mit Osteoarthritis und reaktiver Arthritis verglichen werden. Bei der Durchführung wiederholter Experimente mit den genannten Bedingungen können die Ergebnisse mit dem F-Test verglichen werden, der ebenfalls wie der t-Test eine Normalverteilung voraussetzt und auf einer Varianzanalyse beruht.

Oft ist man auch an Genen interessiert, die einem bestimmten Expressionsprofil folgen und beispielsweise in Synovialbiopsien von Patienten mit RA, aber nicht in anderen Erkrankungen, eine hohe Expression aufweisen. Derartige Fragestellungen können mit Pseudoprofilen analysiert werden, bei denen ein Expressionsmuster vorgegeben wird und die Gene anhand der Übereinstimmung zu diesem vorgegebenen Muster katalogisiert werden (Wu 2001).

Nicht selten führen die aufgeführten statistischen Verfahren zu einer Liste von Kandidatengen, die für die untersuchten Bedingungen möglicherweise funktionell von Bedeutung sind. Im Weiteren muss die differenzielle Expression der Kandidatengene mit unabhängigen Verfahren wie der Real-time-PCR verifiziert und deren funktio-

nelle Bedeutung durch gezielte Veränderung der Expression individueller Gene charakterisiert werden. Hierzu können molekularbiologische Techniken wie die Transfektion mit Expressionsvektoren zur Überexpression eines bestimmten Gens oder dominant-negative Mutanten, Antisense- oder Knockout-Technologien zur Hemmung der Expression eines bestimmten Gens verwendet werden. Da die Anwendung dieser Technologien zeitaufwändig ist, muss zunächst eine genaue Auswahl interessanter Gene erfolgen (Schulze u. Downward 2001). Neuere Verfahren wie die RNA-interference(RNA_i)-Technologie sind möglicherweise auch für die funktionelle Untersuchung einer größeren Anzahl von Genen geeignet (Hammond et al. 2001).

Gruppierung von Genen (Clusterverfahren)

Clusterverfahren sind das am häufigsten angewandte Instrument zur statistischen Analyse von Mikroarrayexperimenten (Brazma u. Vilo 2000). Ziel jeder Clusteranalyse ist es, Objekte so in Gruppen oder Cluster einzuteilen, dass die Objekte eines Clusters einander immer ähnlicher und Objekte aus verschiedenen Clustern einander immer unähnlicher werden. Zur Berechnung der Ähnlichkeit zwischen Objekten werden verschiedene Distanzmaße wie euklidische Distanzen oder deren vielfältige Verallgemeinerungen herangezogen. Auf diese Weise lassen sich beispielsweise Gruppen von Genen bilden, die unter verschiedenen Stimulierungsbedingungen ein ähnliches Expressionsverhalten mit nur geringen, definierten Unterschieden (Distanzen) aufweisen.

Misst man unter Verwendung von Mikroarrays die Expression von Genen in einem Organismus unter verschiedenen Bedingungen, in verschiedenen Entwicklungsstufen oder in verschiedenen Geweben, können die Ergebnisse in einer Expressionsmatrix dargestellt werden (Abb. 1.6). In dieser Expressionsmatrix sind in den Zeilen spezifische Gene zusammengefasst, während die Spalten die verschiedenen Versuchsbedingungen repräsentieren. In einzelnen Feldern ist die Expression eines spezifischen Gens in einer bestimmten Versuchsbedingung ersichtlich. Vergleicht man mit Hilfe von Clusterverfahren die Expressionsprofile in verschiedenen Zeilen der Expressionsmatrix, können auf diese Weise Gruppen von Genen mit gleichem Expressionsverhalten und möglicherweise gleichen Regulationsmechanismen identifiziert werden. Der Vergleich von Spalten in der Expressionsmatrix ermöglicht die Erstellung eines molekularen Profils an differenziell exprimierten Genen in den verschiedenen Versuchsbedingungen. Dieses Verfahren kann beispielsweise bei der Analyse von Gewebeproben auch zur Identifikation von Untergruppen mit einem jeweils spezifischen Expressionsprofil verwendet werden.

Für die Analyse von Mikroarraydaten stehen verschiedene Clusterverfahren zur Verfügung (Brazma u. Vilo 2000). Häufig verwendete Beispiele sind hierarchisch-agglomerative und K-means-Verfahren. Bei hierarchisch-agglomerativen Clusterverfahren wird das hierarchische System durch sukzessive Gruppierung von Genen und im weiteren Verlauf durch Fusion von Gruppen zu größeren Gruppen konstruiert. Die Aggregation beginnt

	Bedingung I	Bedingung II	Bedingung III	Bedingung IV		Bedingung n
Gen 1	x_{11}	x_{12}	x_{13}	x_{14}		x_{1n}
Gen 2	x_{21}	x_{22}	x_{23}	x_{24}		x_{2n}
Gen 3	...	...	...	...		...
Gen 4						
Gen 5						
Gen 6						
Gen 7						
Gen 8						
.						
.						
Gen n						x_{nn}

Abb. 1.6. Expressionsmatrix. In den Zeilen der Expressionsmatrix sind spezifische Gene zusammengefasst, während die Spalten verschiedene Versuchsbedingungen repräsentieren. In den Feldern ist die Expression eines beliebigen Gens x unter einer spezifischen Versuchsbedingung y ablesbar

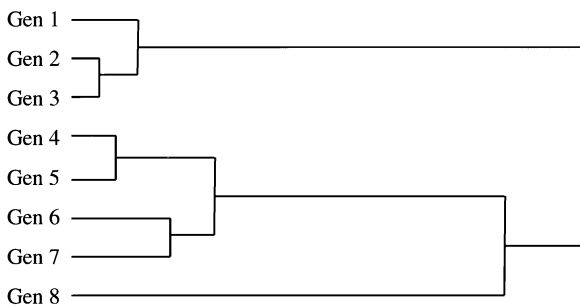


Abb. 1.7. Dendrogramm einer hierarchisch-agglomerativen Clusteranalyse. Je näher der Abstand der neu gebildeten Gruppen zum linken Rand des Dendrogramms, desto homogener ist die Zusammensetzung innerhalb dieser Gruppe. Im aufgeführten Beispiel könnten die Cluster zwischen den Genen 4 und 7 sowie zwischen den Genen 1 und 3 von Interesse sein und auf gemeinsame Regulationsmechanismen innerhalb der Gruppen hinweisen. Alle höher hierarchischen Cluster erscheinen dagegen zu heterogen

hierbei mit der kleinstmöglichen Gruppe, die jeweils nur aus einem Gen besteht. Im weiteren Verlauf werden Gruppen mit dem geringsten Abstand gemäß dem gewählten Distanzmaß in eine neue Gruppe zusammengefasst und anschließend der Abstand zwischen den Gruppen neu errechnet. Hierarchisch-agglomerative Clusteranalysen können als Dendrogramm dargestellt werden (Abb. 1.7). Neben dem hierarchischen System ist in Dendrogrammen auch die Homogenität der gebildeten Gruppen ablesbar. Je früher die Clusterbildung erfolgt, d.h. je näher der Abstand einer neu gebildeten Gruppe vom linken Rand des Dendrogramms ist, desto homogener ist diese Gruppe, d.h. desto ähnlicher sind sich die Gene dieser Gruppe in ihrem Expressionsverhalten. Ein vereinfachtes Beispiel einer hierarchisch-agglomerativen Clusteranalyse ist in Abb. 1.7 dargestellt.

Beim K-means-Clusterverfahren ist im Gegensatz zu den hierarchischen Clusterverfahren die Anzahl der Cluster a priori festgelegt. Anders als bei hierarchischen Verfahren werden auch nicht einzelne, kompakte Cluster auf Kosten hoher Heterogenität in anderen Clustern gebildet. Stattdessen wird beim K-means-Verfahren nach Gruppierungen gesucht, bei denen jedes Objekt (z.B. die Expression eines bestimmten Gens) zum Schwerpunkt seiner Gruppe einen kleineren Abstand besitzt als zu den anderen Gruppenschwerpunkten. Auf diese Weise entstehen Gruppen mit mittlerer Homogenität, während die Bildung heterogener Gruppen vermieden wird.

Clusterverfahren sind wiederholt erfolgreich zur Analyse von Mikroarraydaten eingesetzt worden.

Eine interessante Anwendung ist die Identifikation von Tumorsubklassen durch hierarchische Clusterverfahren. Alidazeh et al. (2000) untersuchten die molekularen Expressionsmuster von Patienten mit diffusem großzelligem B-Zell-Lymphom (DLBCL), dem häufigsten Subtyp der Non-Hodgkin-Lymphome. Der klinische Verlauf dieser Erkrankung ist äußerst heterogen. Zwar spricht initial die Mehrzahl der Patienten auf die Chemotherapie an, allerdings erreichen nur weniger als die Hälfte der Patienten eine dauerhafte Remission mit entsprechend besserer Lebenserwartung. Für die Erstellung eines Expressionsprofils wurden 96 Proben aus gesunden und malignen Lymphozyten untersucht. Unter Verwendung eines hierarchischen Clusteralgorithmus konnte gezeigt werden, dass 2 unterschiedliche Subgruppen von DLBCL-Lymphomen mit jeweils spezifischem Expressionsmuster existieren. Die 2 anhand ihres Expressionsprofils definierten DLBCL-Untergruppen korrelierten gut mit der Lebenserwartung der Patienten. Die molekulare Klassifikation von Tumoren auf der Basis von Expressionsprofilen kann auf diese Weise zur Identifikation bisher unbekannter, aber klinisch signifikanter Tumorsubtypen eingesetzt werden.

1.3.3.5 Speicherung der Ergebnisse in Datenbanken

Die großen Mengen an Daten aus Mikroarrayexperimenten sind für herkömmliche Formen wissenschaftlicher Publikationen nur bedingt geeignet. Wissenschaftliche Zeitschriften versuchen das Problem limitierter Seitenzahlen einerseits und großer Datenmengen andererseits durch Veröffentlichungen von Anteilen der Daten als „zusätzliche Information“ auf ihren Webseiten zu lösen. Ein Problem dieser Publikationsform besteht darin, dass die Daten durch Veröffentlichung in verschiedensten Zeitschriften in nicht standardisierten Formaten vorliegen. Hierdurch werden die Aussagekraft und Vergleichbarkeit der Daten erheblich eingeschränkt, da insbesondere Mikroarrayexperimente stark durch die Art des Arrays, die experimentellen Bedingungen, die Form der Datenerfassung und Auswertung und zahlreicher anderer Parameter beeinflusst werden.

Aus diesen Gründen bestehen derzeit Bestrebungen, Daten von Mikroarrayexperimenten in öffentlich zugänglichen Datenbanken nach einem standardisierten Erfassungsprotokoll zu veröffentlichen (Brazma et al. 2001). Als Vorbild können die GenBank und andere Datenbanken dienen, die zur Speicherung von Sequenzierdaten nach einem

standardisierten Verfahren benutzt werden. Richtlinien für ein standardisiertes Protokoll zur Erfassung von Mindestinformationen über Mikroarray-experimente (MIAME, minimum information about microarray experiments) wurden von der Microarray-Gene-Expression-Database-Gruppe entwickelt (<http://www.mged.org>) (Brazma et al. 2001). Nach diesen Richtlinien sollte ein MIAME-Dokument Informationen über folgende Bereiche enthalten:

1. Experimentelle Bedingungen
 - Art des Experiments mit getesteten Bedingungen und gewählten Parametern,
 - Art und Anzahl der Wiederholungen,
 - Qualitätskontrollen
2. Design des Arrays
 - Beschreibung der Art des Arrays inklusive Hersteller sowie Informationen über die Art der DNA-Spots (Oligonukleotide/cDNA),
 - Sequenzierinformation zu jedem Spot,
 - Kontrollspots
3. Proben
 - Herkunft der Proben (Zellart, Organismus, Gewebe),
 - Art der RNA-Extraktion,
 - Markierung der Proben
4. Hybridisierung
 - Art der Hybridisierungslösung,
 - Waschvorgänge,
 - Menge der markierten Probe,
 - Hybridisierungsdauer,
 - Volumen,
 - Temperatur und Art der Hybridisierungskammer
5. Messungen
 - Speicherung des Rohdatenbilds als Image-File (z.B. TIFF),
 - Quantifizierungsmatrix mit Informationen über Bildverarbeitungssoftware und Art der Quantifizierung,
 - Expressionsmatrix
6. Normalisierung
 - Normalisierungsverfahren
 - Art der Normalisierungskontrollen

Um automatisierte Suchfunktionen zu erleichtern, soll die Dokumentation in MIAME im Gegensatz zu herkömmlichen Publikationen nicht als freier Text, sondern strukturiert mit definiertem Vokabular erfolgen. Die Einrichtung entsprechender Datenbanken ist derzeit durch verschiedene Institutionen (National Center for Biotechnology Information, DNA Database of Japan, European Bioinformatics Institute) in der Entwicklung.

1.3.4 Anwendungsbeispiel in der Rheumatologie

In den letzten Jahren konnten vermehrt Hinweise gefunden werden, dass T-Zell-unabhängige Stoffwechselwege entscheidend an der progressiven Gelenkzerstörung der RA beteiligt sind. Einen wichtigen Beleg hierfür lieferte das SCID-Maus-Modell der RA, bei dem humane synoviale Fibroblasten zusammen mit humanem Knorpel unter die Nierenkapsel von SCID-Mäusen gebracht wurden (Muller-Ladner et al. 1996). In diesem Modell kann nach 60 Tagen die Invasion der synovialen Fibroblasten in den Knorpel semiquantitativ analysiert werden. Dabei zeigen synoviale Fibroblasten von Patienten mit RA eine deutliche Invasion in den Knorpel, während Kontrollfibroblasten keine oder nur eine geringfügige Invasion aufweisen. Da SCID-Mäuse einen Defekt der T-Zellen aufweisen, deuten diese Ergebnisse darauf hin, dass die Knorpeldestruktion der RA durch intrinsische Eigenschaften der rheumatoiden synovialen Fibroblasten entscheidend mitbeeinflusst wird.

Zur Identifikation von Molekülen, die für die destruierenden Eigenschaften der rheumatoiden Fibroblasten verantwortlich sein könnten, wurde eine Kombination von Differential-display und Mikroarray angewandt (Neumann et al. 2002). Hierfür wurde zunächst RNA von kultivierten synovialen Fibroblasten von Patienten mit RA sowie OA extrahiert, im Rahmen der reversen Transkription radioaktiv markiert und mit einem kommerziellen cDNA-Array auf Nylonmembranbasis hybridisiert. Nachteile dieses klassischen Verfahrens bestehen in der relativ geringen Sensitivität der Methode, welche die Detektion von Molekülen mit geringer Expression erschwert, und in den relativ großen Mengen an RNA, die für die Experimente benötigt werden. Aus diesen Gründen wurde dem Array eine RNA-arbitrarily-primed-Polymerasekettenreaktion (RAP-PCR) vorgeschaltet. Bei diesem Differential-display-Verfahren werden in den PCR-Schritten arbiträre Primer verwendet, wodurch es zu einer Amplifizierung der RNA kommt.

Durch die Kombination der beiden Verfahren konnte bei etwa 50% der Gene auf dem Array ein reproduzierbares Hybridisierungssignal detektiert werden, während die Anzahl detektierbarer Gene mit der klassischen Methode deutlich geringer war (Neumann et al. 2002). Mit der kombinierten Methode konnten 8 Gene identifiziert werden, die bei RA-Fibroblasten höher als bei OA-Fibroblasten exprimiert waren. 4 Gene zeigten ein umgekehrtes Expressionsmuster. Exemplarisch wurde mittels Real-time-PCR sowie In-situ-Hybridisierung die

erhöhte Expression des CD82-Antigens in RA-Fibroblasten verifiziert. Bei dem CD82-Antigen handelt es sich um ein multifunktionelles Tetraspanin, das bisher nicht mit der Pathogenese der RA assoziiert wurde.

Das kombinierte RAP-PCR-Mikroarray-Verfahren erlaubt die Arbeit mit sehr geringen RNA-Mengen zur Identifikation von differenziell exprimierten Genen. Durch die Verwendung von arbiträren Primern erfolgt allerdings eine selektive und unvollständige Amplifizierung der RNA-Population, sodass sich dieses Verfahren nicht zur Erstellung eines Expressionsprofils eignet.

1.4 Resümee

In der rheumatologischen Grundlagenforschung wurden in den letzten Jahren zahlreiche Zytokine, Adhäsionsmoleküle, Wachstumsfaktoren, Regulatoren der Apoptose und Matrix zerstörende Enzyme beschrieben. Weniger bekannt, aber für die Identifikation neuer therapeutischer Zielmoleküle von entscheidender Bedeutung, sind dagegen die Auswirkungen dieser Moleküle auf die Expression anderer Gene. Die suppressive subtraktive Hybridisierung (SSH) und Mikroarrays stellen Verfahren zur Identifikation differenziell exprimierter Gene dar. Bei der SSH besteht das Prinzip in einer Hybridisierung von Sequenzen, die sowohl in der Testprobe als auch in der Vergleichsprobe vorhanden sind. Sequenzen ohne Hybridisierungspartner repräsentieren differenziell exprimierte Gene, die durch anschließende PCR amplifiziert werden können. Die SSH eignet sich insbesondere zur Identifikation von Sequenzen mit noch unbekannter biologischer Funktion. Zur Erstellung eines kompletten Expressionsprofils ist dieses Verfahren aber zu zeit- und zu kostenintensiv.

Die Anwendungsmöglichkeiten von Mikroarrays sind vielfältiger als diejenigen der SSH. Mikroarrays können u. a. zur Identifikation von Kandidatengen, zur Suche nach Krankheitssubklassen, zum Aufsuchen gemeinsamer molekularer Regulationsfaktoren durch vergleichbares Expressionsverhalten und zum Screening nach Polymorphismen eingesetzt werden.

Klassische wissenschaftliche Experimente beruhen auf wissenschaftlichen Hypothesen, die durch die Ergebnisse der Experimente bewiesen oder widerlegt werden. Im Gegensatz dazu setzen Mikroarrayexperimente nicht zwingend eine wissen-

schaftliche Hypothese voraus. Vielmehr können Mikroarrays auch Hypothesen generierende Funktionen haben, indem beispielsweise neue Krankheitssubklassen einer zuvor uniformen Entität postuliert werden. Es ist wichtig zu betonen, dass diese Hypothesen generierende Funktion selbstverständlich nicht von der Notwendigkeit befreit, die neu postulierte Hypothese mit klassischen Methoden zu verifizieren.

Viele Autoren sehen eine Hauptgefahr der Mikroarrays darin, dass der Reiz der neuen Technologie viele Wissenschaftler zu einer kritiklosen Anwendung verleiten könnte, ohne sich der Limitierungen der Methodik ausreichend bewusst zu sein. Im rheumatologischen Bereich wurden daher Richtlinien für Studien mit Mikroarrays erlassen (Firestein u. Pisetsky 2002). Diese Richtlinien umfassen die Identifikation falsch-positiver Ergebnisse inklusive einer Untersuchung der Variabilität spezifischer Gene in verschiedenen Experimenten, die statistische Auswertung verbunden mit einer adäquaten Anzahl an Wiederholungen sowie die Bestätigung der differentiellen Expression von Schlüssel-molekülen mit unabhängigen Methoden. Werden diese und ähnliche Richtlinien befolgt, stellen Mikroarrays einen wichtigen technischen Fortschritt dar, der die molekularbiologische Forschung in den nächsten Jahren rasant beschleunigen kann.

1.5 Literatur

- Alizadeh AA, Eisen MB, Davis RE et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503–511
- Bowtell DD (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nat Genet* 21: 25–32
- Brazma A, Vilo J (2000) Gene expression data analysis. *FEBS Lett* 480: 17–24
- Brazma A, Hingamp P, Quackenbush J et al. (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* 29: 365–371
- Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G (1999) Making and reading microarrays. *Nat Genet* 21:15–19
- Churchill GA, Oliver B (2001) Sex, flies and microarrays. *Nat Genet* 29: 355–356
- Desaj S, Hill J, Trelogan S, Diatchenko L, Siebert PD (2000) Identification of differentially expressed genes by suppression subtractive hybridization. In: Hunt SP, Livesey FJ (eds) *Functional genomics*. Oxford University Press, Oxford New York, pp 81–112
- Diatchenko L, Lau YF, Campbell AP et al. (1996) Suppression subtractive hybridization: a method for generating

- differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93: 6025–6030
- Distler J, Hirth A, Scheid A et al. (2002) Gene profiling of scleroderma fibroblasts after exposure to hypoxia. *Ann Rheum Dis* 61
- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* 21: 10–14
- Firestein GS, Pisetsky DS (2002) DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology. *Arthritis Rheum* 46: 859–861
- Frost, Sullivan (2001) The World biochip market. Report 7761, <http://www.frost.com>
- Hammond SM, Caudy AA, Hannon GJ (2001) Post-transcriptional gene silencing by double-stranded RNA. *Nat Rev Genet* 2: 110–119
- Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Res* 6: 986–994
- Hieter P, Boguski M (1997) Functional genomics: it's all how you read it. *Science* 278: 601–602
- Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci USA* 88: 7276–7280
- Knight J (2001) When the chips are down. *Nature* 410: 860–861
- Lechner S, Muller-Ladner U, Neumann E et al. (2001) Use of simplified transcripts for the analysis of gene expression profiles in laser-microdissected cell populations. *Lab Invest* 81: 1233–1242
- Li X, Gu W, Mohan S, Baylink DJ (2002) DNA microarrays: their use and misuse. *Microcirculation* 9: 13–22
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* 21: 20–24
- Lockhart DJ, Winzler EA (2000) Genomics, gene expression and DNA arrays. *Nature* 405: 827–836
- Lukyanov KA, Launer GA, Tarabykin VS, Zaraisky AG, Lukyanov SA (1995) Inverted terminal repeats permit the average length of amplified DNA fragments to be regulated during preparation of cDNA libraries by polymerase chain reaction. *Anal Biochem* 229: 198–202
- McPherson JD, Marra M, Hillier L et al. (2001) A physical map of the human genome. *Nature* 409: 934–941
- Mills JC, Roth KA, Cagan RL, Gordon JI (2001) DNA microarrays and beyond: completing the journey from tissue to cell. *Nat Cell Biol* 3: E175–E178
- Mousses S, Bittner ML, Chen Y et al. (2000) Gene expression analysis by cDNA microarray. In: Hunt SP, Livesey FJ (eds) *Functional genomics*. Oxford Press, Oxford New York, pp 113–138
- Mueller BM, Yu YB, Laug WE (1995) Overexpression of plasminogen activator inhibitor 2 in human melanoma cells inhibits spontaneous metastasis in scid/scid mice. *Proc Natl Acad Sci USA* 92: 205–209
- Muller-Ladner U, Kriegsmann J, Franklin BN et al. (1996) Synovial fibroblasts of patients with rheumatoid arthritis attach to and invade normal human cartilage when engrafted into SCID mice. *Am J Pathol* 149: 1607–1615
- Neidhart M, Rethage J, Kuchen S et al. (2000) Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum* 43: 2634–2647
- Neumann E, Kullmann F, Judex M et al. (2002) Identification of differentially expressed genes in rheumatoid arthritis by a combination of complementary DNA array and RNA arbitrarily primed-polymerase chain reaction. *Arthritis Rheum* 46: 52–63
- Schulze A, Downward J (2001) Navigating gene expression using microarrays – a technology review. *Nat Cell Biol* 3: E190–E195
- Southern E, Mir K, Shchepinov M (1999) Molecular interactions on microarrays. *Nat Genet* 21: 5–9
- Stein OD von, Thies WG, Hofmann M (1997) A high throughput screening for rarely transcribed differentially expressed genes. *Nucleic Acids Res* 25: 2598–2602
- Vandesompele J, De Paepe A, Speleman F (2002) Elimination of primer-dimer artifacts and genomic coamplification using a two-step SYBR green I real-time RT-PCR. *Anal Biochem* 303: 95–98
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* 270: 484–487
- Venter JC, Adams MD, Myers EW et al. (2001) The sequence of the human genome. *Science* 291: 1304–1351
- Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M (1992) Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* 20: 4965–4970
- Woo TH, Patel BK, Cinco M et al. (1998) Real-time homogeneous assay of rapid cycle polymerase chain reaction product for identification of *Leptonema illini*. *Anal Biochem* 259: 112–117
- Wu TD (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J Pathol* 195: 53–65
- Zhang J, Underwood LE, D'Ercole AJ (2001) Hepatic mRNAs up-regulated by starvation: an expression profile determined by suppression subtractive hybridization. *FASEB J* 15: 1261–1263
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30: 892–897

Molekularmedizinische Grundlagen von rheumatischen
Erkrankungen

Ganten, D.; Ruckpaul, K.; Gay, S.; Kalden, J.R. (Hrsg.)

2003, XX, 394 S. 476 Abb., 440 Abb. in Farbe.,

Hardcover

ISBN: 978-3-540-43735-2