

1. Einleitung

Die vorliegende Arbeit ist Teil eines Projektes zur Aufklärung von Mechanismen der Neuroprotektion und Ischämietoleranz bei cerebraler Ischämie. Ein wichtiger Bestandteil zum näheren Verständnis der zugrundeliegenden Mechanismen ist dabei die Analyse der Genexpression nach cerebraler Ischämie als Korrelat dieses klinisch relevanten Phänomens.

1.1. Der Schlaganfall

Der Schlaganfall stellt mit einer Inzidenz von 200-350/100000 eine der häufigsten Krankheitsgruppen und zugleich die dritthäufigste Todesursache in den Industrienationen dar (*Murray et al 1997*). In 80-85 % der Fälle liegen fokale, ischämisch bedingte Durchblutungsstörungen zugrunde. Aus sozioökonomischer Sicht wird die Bedeutung des Schlaganfalls durch folgende Zahlen ersichtlich: Nur 1/3 der überlebenden Patienten kann nach dem Ereignis wieder eine geregelte Tätigkeit aufnehmen, 1/3 wird berufs- und erwerbsunfähig, 1/3 der Patienten bleibt dauerhaft pflegebedürftig. Initial versterben bis zu etwa 30 % der Schlaganfallpatienten innerhalb der ersten vier Wochen (*Bronnum-Hansen et al 2001*) und jeder zweite überlebende Patient erleidet innerhalb von 5 Jahren ein Rezidiv oder einen Herzinfarkt an dem er verstirbt.

Die wesentlichen Therapiestrategien des ischämischen Schlaganfalls bestehen in Maßnahmen zur Blutdruck- und Blutzuckereinstellung, Fiebersenkung, Überwachung der Atemfunktion und Oxygenierung, Hirnödembehandlung, Thromboseprophylaxe, ggf. Thrombolysetherapie mit rt-PA und Einleitung einer Sekundärprophylaxe.

Zusätzlich wird seit vielen Jahren intensiv nach neuroprotektiven Therapien geforscht, die die Ischämietoleranz der Neuronen verbessern können, vor weiterer Zellschädigung nach dem ischämischen Ereignis schützen und den Prozess der Infarktprogression in der Penumbra stoppen sollen.

1.2. Tiermodell der fokalen cerebralen Ischämie

Ein häufig angewendetes Tiermodell zur Untersuchung der Pathophysiologie der cerebralen Ischämie ist die fokale transiente (oder permanente) Okklusion der A. cerebri media (MCA) in der Maus oder Ratte (*Hara et al 1996*).

Das dabei entstehende Infarktgebiet kann in einen Kern und in eine Penumbra eingeteilt werden. Im Kern beträgt der cerebrale Blutfluss weniger als 10 % der Ausgangsrate, um den Kern findet man eine schwächer ischämische Zone, die Penumbra, mit Blutflusswerten um

30-35 % des Ausgangsniveaus (*Hossmann 1994*). Der Infarktkern stellt irreversibel geschädigtes, später nekrotisches Gewebe dar. Die ischämische Penumbra ist ein funktionsgestörtes Areal mit noch reversiblen strukturellen Schäden, die jedoch eine Progression zum Infarkt zeigen können.

1.3. Genexpression nach fokaler cerebraler Ischämie

Der Gewebeschaden nach cerebraler Ischämie resultiert aus der Interaktion einer Vielzahl komplexer pathophysiologischer Prozesse wie Periinfarkt-Depolarisierung, Exzitotoxizität, Inflammation und Apoptose (*Dirnagl et al 1999*). In deren Rahmen wird eine Fülle von Genexpressionsänderungen im Gehirn induziert, deren Einfluss sehr komplex und zum Teil noch ungeklärt sind.

Für fast alle Komponenten der ischämischen Kaskade konnten tierexperimentell pharmakologische Ansatzpunkte zur neuroprotektiven Intervention gefunden werden, darunter auch gentherapeutische oder molekularbiologisch basierte Therapieansätze wie die Verwendung transgener Tiere, viraler Transfer neuroprotektiver Gene und die Applikation von Antisense mRNA zur Unterdrückung der Translation destruktiver Gene. Die meisten dieser Ansätze greifen in die Kaskade der postischämischen Inflammation und des (verzögerten) Zelltodes ein.

Bereits wenige Minuten nach fokaler cerebraler Ischämie kommt es inner- und außerhalb des Infarktgebietes zur transienten Induktion sog. *immediate early genes*, von denen viele für DNA-bindende Transkriptionsfaktoren, wie z.B. AP-1, NFκB oder p53 kodieren. Diese regulieren die Expression von zahlreichen späten Effektorgenen. So konnte in einem Ratten-Schlaganfallmodell mit reversibler fokaler Ischämie bereits innerhalb einer Stunde eine Induktion von Proteinen der jun- und fos-Familie detektiert werden (*An et al 1993, Hata et al 2000*). Die Expression der *immediate early genes* steht in engem Zusammenhang mit der postischämisch ausgelösten inflammatorischen Reaktion. Die *immediate early genes* wirken transaktivierend (z.B. über NFκB) und führen zu einer frühen (1-3 h) Induktion von Zytokinen wie TNF-alpha (*Barone et al 1997*), IL-1β (*Zhang et al 1998*), IL-6 (*Suzuki et al 1999*), IL-10 (*Zhai et al 1997*) in geschädigten Neuronen, Astrozyten, Glia und Endothelzellen. Über diese werden chemotaktische Signale, Apoptose und Zelltod sowie auch protektive Effekte vermittelt. In überlebenden Gefäßen des Infarktzentrums erfolgt die Induktion und Expression zahlreicher Gene, die die Invasion inflammatorischer Zellen in das Infarktgebiet fördern, z.B. Matrix Metalloproteinasen (*Gasche et al 1999, Heo et al 1999*) und Adhäsionsmoleküle wie ICAM-1, ELAM-1, P- und E-Selektine (*Del Zoppo et al 2000*).

Verschiedene Studien in experimentellen Schlaganfall-Modellen konnten zeigen, dass eine Hemmung der postischämischen inflammatorischen Reaktion durch die Inhibition von Adhäsionsmolekülen (*Shibata et al 2001*), die Blockade von IL-1 (*Loddick et al 1996*), anti TNF-Ak oder Liganden (*Yang 1998*), bzw. die Applikation von IL-6 (*Loddick et al 1998*) oder iNOS Inhibition (*Parmentier et al 2001*) den ischämischen Schaden verringert.

Innerhalb des gleichen Zeitfensters wie die *immediate early genes* findet sich im Infarktkern und in der Penumbra eine Induktion zahlreicher Hitzeschock-Protein-mRNAs, darunter z.B. Hsp70 (*Kinouchi et al 1993*), HSC70 (*Abe et al 1993*) oder Hsp32 (*Nimura et al 1996*).

In der Zone um den Infarktkern werden in Abhängigkeit vom Überleben oder Untergang der Zellen durch DNA-Schädigung induzierbare Gene, DNA-Reparaturgene, pro- und antiapoptotische Gene, Zell-Zyklus-Gene und Caspasen induziert. In letal geschädigten Zellen des Infarktzentrums sowie im Randbereich findet sich z.B. eine Hochregulation von Bax, p53, MDM2 und Gadd 45 (*Li et al 1997*). Zusätzlich läßt sich eine Induktion von Caspasen nachweisen, wobei insbesondere den Caspasen 1 und 3 eine wesentliche Rolle beim ischämievermittelten Zelltod durch Apoptose und Nekrose zugeschrieben wird (*Dirnagl et al 1999*). Bcl-2 und Bcl-xl, zwei antiapoptotische Gene werden in wahrscheinlich überlebenden Zellen des Randgebietes induziert (*Chen et al 1995, Isenmann et al 1998*). Im Modell der fokalen cerebralen Ischämie der Ratte führte die Expression von bcl-2 mittels eines cortical injizierten HSV1-Vektors zu reduziertem neuronalen Tod (*Lawrence et al 1996*). Transgene bcl-2 überexprimierende Mäuse entwickelten nach fokaler cerebraler Ischämie ein geringeres Infarkt volumen (*Martinou et al 1994*). Eine Inhibition des neuronalen Zelltodes nach cerebraler Ischämie konnte auch mit Caspaseinhibitoren nachgewiesen werden (*Endres et al 1998*).

Zeitlich verzögert läßt sich im Ischämiegebiet die Expression diverser Wachstumsfaktoren und deren Rezeptoren wie z.B. TGF-beta1 (*Lehrmann et al 1998*), bFGF (*Lin et al 1997, Speliotos et al 1996*), NGF (*Hsu et al 1993*), BDNF (*Hsu et al 1993*), *glial derived nerve growth factor* (*Arvidsson et al 2001*), *insulin growth factor* und bindende Proteine (*Gluckmann et al 1992*), *platelet derived growth factor* (*Iihara et al 1994*), *ciliary neurotrophic factor* (*Lin et al 1998*), *growth inhibitory factor* (Metallothionein III) (*Yuguchi et al 1997*) nachgewiesen werden. Für viele dieser neurotrophen Faktoren wurde im Tiermodell eine protektive Wirkung gegen fokale Ischämie nachgewiesen (*Semkova et al 1999*).

1.4. Detektion differentiell exprimierter Gene nach fokaler cerebraler Ischämie

Während tierexperimentell mit vielen Substanzen große Fortschritte im Hinblick auf die Reduktion der Infarktgröße und des resultierenden neurologischen Defizits erzielt werden konnten, blieben klinische Studien mit neuroprotektiven Medikamenten wie z.B. Natrium- und Calciumkanal-Antagonisten, NMDA-Rezeptor-Antagonisten, GABA-Agonisten, freie Radikalfänger, NO-Pathway-Modulatoren, Blocker von Adhäsionsmolekülen bisher erfolglos (*De Keyser et al 1999, Fisher et al 2000*). Trotz der unbefriedigenden Ergebnisse bisheriger klinischer Studien mit neuroprotektiven Medikamenten besteht weiterhin die Hoffnung, klinisch einsetzbare neuroprotektive Substanzen zur Behandlung des akuten Schlaganfalls und zur Förderung der Regeneration zu finden. Eine Möglichkeit besteht darin, auf molekularer Ebene nach zeitabhängig spezifischen Genen, die im Zusammenhang mit dem ischämischen Schaden und dem Regenerationsprozess nach einem Schlaganfall stehen, zu suchen. Das Verständnis der transkriptionellen Änderungen im Gehirn nach der Ischämie kann Informationen über die Mechanismen, die zur induzierten Ischämietoleranz im Gehirn führen bzw. ihr Korrelat darstellen, liefern.

Ein besonderes Interesse liegt dabei in der Identifizierung noch unbekannter neuroprotektiver Gene, die potentielle therapeutische Ziele darstellen könnten.

Grundsätzlich kann man zur Analyse der Genexpression einem Kandidatenansatz (in situ Hybridisierung, RT-PCR, Northern Blot) oder einem Screening-Ansatz (Differential Display, Arrays, subtraktive PCR-/Hybridisierungsmethoden, SAGE) folgen. Der Screening-Ansatz ermöglicht die Identifikation von Gruppen differentiell exprimierter bekannter und unbekannter Gene, während der Kandidatenansatz die selektive Auswahl spezifischer interessanter Gene beinhaltet.

Mittels Kandidatenansatz konnten bereits viele ischämierrelevante Gene identifiziert werden, darunter z.B. Stanniocalcin (*Zhang et al 2000*) und Metallothionein-1 (*van Lookeren Campagne et al 1999*).

Im Gegensatz dazu, sollen Screening-Methoden durch die globale Analyse des Transkriptom ein umfassenderes Verständnis der postischämischen Genexpression ermöglichen und Möglichkeiten zur Detektion neuer (bekannter und unbekannter) ischämierrelevanter Gene eröffnen.

So wurden in einem Oligonuklotid-Array am Periinfarkt-Cortex der Ratte nach permanenter fokaler Ischämie unter 328 differentiell regulierten Transkripten etwa 50 % gefunden, die vorher noch nicht als reguliert beschrieben waren (*Lu et al 2003*).

Schwarz und Mitarbeiter analysierten mit einer subtraktiven PCR-Methode die Genexpression nach transienter fokaler Ischämie in der Ratte und konnten vier bisher noch nicht beschriebene differentiell exprimierte Gene (Neurales Pentraxin 2, bcl-2 bindendes Protein BIS, Cystein reiches Protein 61, Lektin ähnlicher ox-LDL Rezeptor) detektieren (*Schwarz et al 2002*).

Wang und Mitarbeiter isolierten mittels Differential Display nach fokaler Okklusion der A. cerebri media in der Ratte Adrenomedullin, welches sich als ischämie-relevantes destruktives Gen erwies (*Wang et al 1995*).

Ein weiteres, erst 1995 von Velculescu (*Velculescu et al 1995*) entwickeltes Verfahren zur globalen Analyse des Genexpressionsprofils stellt die Serielle Analyse der Genexpression (SAGE) dar. Ein wesentlicher Vorteil der SAGE-Technik gegenüber anderen Screening-Methoden zur Genexpressionsanalyse liegt in der Möglichkeit zur Detektion neuer oder bisher wenig charakterisierter differentiell exprimierter Gene. Hier eröffnen sich auch für die Schlaganfallforschung große Potentiale zur Identifikation neuer neuroprotektiver Gene.

1.5. Das Prinzip der SAGE-Technik (Serial Analysis of Gene Expression)

Die von Velculescu 1995 (*Velculescu et al 1995*) erstmalig beschriebene SAGE-Technik ist eine Methode zur qualitativen und quantitativen Untersuchung der Genexpressionsprofile von Zellen und Geweben.

Sie basiert auf zwei Prinzipien: 1.) eine isolierte kurze 9-10 bp Nukleotidsequenz (*tag*) von einer definierten Stelle in einem Transkript (mRNA) enthält ausreichend Information, um dieses Transkript eindeutig zu identifizieren, 2.) durch die Verknüpfung dieser kurzen Sequenz-*tags* und anschließende Klonierung wird eine effiziente Analyse von gleichzeitig vielen *tags* durch Sequenzierung eines einzelnen Klons ermöglicht.

Die Generierung einer SAGE-cDNA Expressionsbibliothek beinhaltet die folgenden Schritte: Aus der isolierten mRNA von Zellen oder Geweben wird mit Hilfe eines biotinylierten Oligo(dT) Primers eine Doppelstrang cDNA synthetisiert (Abb. 1A). Anschließend wird die cDNA mit einem Restriktionsenzym (sog. *anchoring* Enzym, Abb. 1B), welches die meisten Transkripte mindestens einmal schneidet, verdaut. In der ursprünglich beschriebenen Methode wird dafür NlaIII verwendet, jedoch sind inzwischen auch andere Enzyme z.B. Sau3AI (*Virlon et al 1999*) beschrieben worden. Typischerweise werden dafür Restriktionsenzyme verwendet, die eine Erkennungssequenz von 4 bp besitzen, weil diese im Durchschnitt alle 4⁴ (256) bp schneiden, während die meisten Transkripte wesentlich länger sind. Nach dem

Restriktionsverdau erfolgt die Isolation der 3'-terminalen cDNA-Fragmente durch Bindung an Streptavidin beschichtete paramagnetische Kügelchen (*streptavidin beads*) über den biotinylierten Oligo(dT) Primer (Abb. 1B). Im nächsten Schritt wird die gebundene cDNA in zwei Fraktionen geteilt und je einer von zwei verschiedenen Linkern an die Restriktionsstellen ligiert (Abb. 1C). Diese Linker enthalten eine Typ IIS Restriktionsschnittstelle. Restriktionsendonukleasen vom Typ IIS (*Pingoud et al 2001*) schneiden an einer sequenzunabhängigen definierten Entfernung bis zu 20 bp von ihrer asymmetrischen Erkennungssequenz. Die Linker sind so konstruiert, dass der Verdau der gebundenen cDNA mit einem derartigen Restriktionsenzym (sog. *tagging* Enzym, z.B. BsmFI) zu einem Stück führt, welches aus Linker und einer kurzen transkriptspezifischen cDNA-Sequenz (*tag*) besteht (Abb. 1D). Nach dem Verdau mit dem *tagging* Enzym werden die freigesetzten Sequenzstücke beider cDNA Fraktionen gegeneinander ligiert, es entstehen sogenannte *ditags*. Es wird dabei davon ausgegangen, dass die Chance für zwei gleiche *tags* in einem *ditag* gekoppelt zu werden, selbst für häufig vorkommende *tags* gering ist. Deshalb werden aus zwei gleichen *tags* bestehende *ditags*, wenn sie bei der späteren Analyse auftreten, als Artefakt ausgeschlossen. Die *ditags* werden anschließend mit linkerspezifischen Primern in einer PCR amplifiziert (Abb. 1E). Danach werden die PCR-Produkte mit dem zuvor verwendeten *anchoring* Enzym (z.B. NlaIII) restriktionsverdaut (Abb. 1F). Dadurch erfolgt die Trennung der *ditags* von den Linkern. Die isolierten *ditags* werden durch Ligation miteinander verknüpft. Die entstandenen Concatemere werden kloniert und anschließend sequenziert. Danach erfolgt die Auszählung der spezifischen *tags*. Es wird bestimmt, wie oft jedes einzelne *tag* auftritt. Zum Ausschluß von PCR-Artefakten wird jede *ditag*-Kombination nur einmal gezählt. Weiterhin werden *ditag*-Kombinationen mit Linkersequenzen oder zwei identischen *tags* eliminiert.

SAGE führt zu einer quantitativen Expressionsaussage in digitaler Form. Es wird im Gegensatz zu anderen Methoden wie Differential Display oder subtraktiven Methoden das gesamte Expressionsprofil der untersuchten Zellen und Gewebe analysiert. SAGE ist dabei ca. 25-30-fach effizienter als die konventionelle EST-Analyse (*Velculescu et al 1995*). SAGE ermöglicht neben der Detektion neuer Gene (*gene discovery*) auch die Detektion von verschiedenen Spleißvarianten eines Gens.

Die grundlegende Frage die sich bei der Auswertung der SAGE Daten stellt, ist die Differenz der *tag*-Anzahl zwischen zwei SAGE-cDNA-Bibliotheken, welche eine zuverlässige Schlussfolgerung über eine differentielle Expression erlaubt (*Audic et al 1997*). Zur statistischen Auswertung von SAGE Experimenten finden unter anderem der Chi-Quadrat

Test, der Fishers Exact Test und die Claveries Bayesian Method (*Chen B H et al 1998*) Anwendung (*Man et al 2000*).

Inzwischen wurden vielfältig modifizierte SAGE-Protokolle zur Optimierung und Effizienzsteigerung beschrieben: z.B. durch Reduktion der Ausgangsmenge der Gesamt-RNA (*Datson et al 1999, Peters et al 1999, Virlon et al 1999, Neilson 2000, Ye et al 2000, Tremain et al 2001*), durch Erhöhung der *ditag*-Ausbeute (*Angelaastro et al 2000, Mathupala et al 2002*) oder durch Erhöhung der Zahl der *tags* je Klon (*Powell 1998, Kenzelmann et al 1999*).

Serial Analysis of Gene Expression (SAGE)

V. E. Velculescu et al., Science 1995

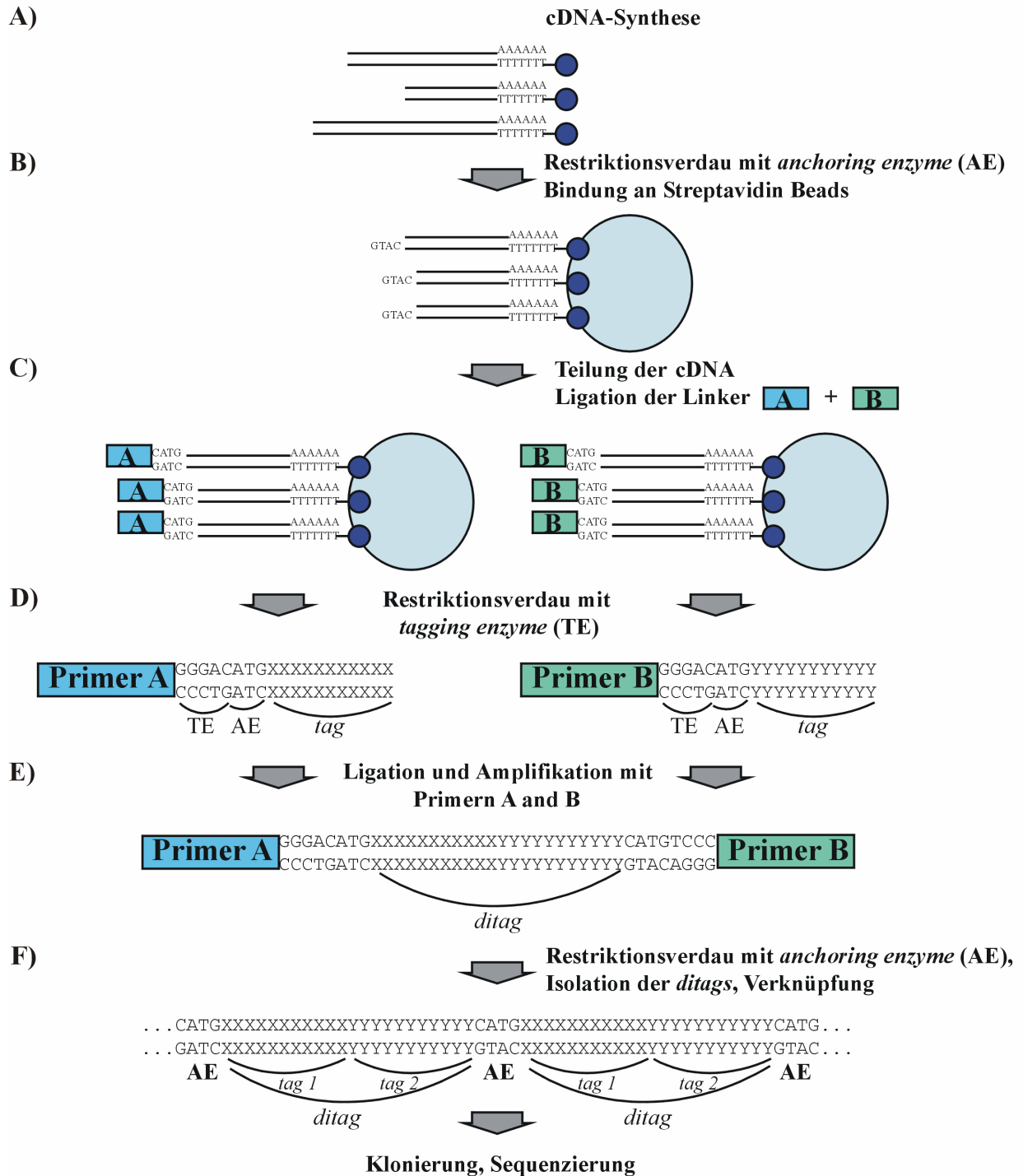


Abb. 1: Prinzip der SAGE-Technik

Vorraussetzung für die weitere Analyse der SAGE-Daten ist die Zuordnung der einzelnen SAGE-tags zu ihrem korrespondierenden Gen.

1.5.1. Methoden zur SAGE *tag*-spezifischen Transkriptidentifikation

Eine herkömmliche Methode zur Identifizierung SAGE-*tag* korrespondierender Transkripte besteht in der Homologierecherche in den zur Verfügung stehenden Referenzsequenz-Datenbanken. Diese kann automatisiert über das *SAGEmap* Programm „*tag to gene mapping*“ des NCBI (<http://www.ncbi.nlm.nih.gov/SAGE>) oder manuell mit Hilfe des BLAST-Algorithmus (<http://www.ncbi.nih.gov/BLAST>) erfolgen.

Die Grundlage für die automatisierte Zuordnung von SAGE-*tags* über das *SAGEmap* Programm bilden den Sequenz-Clustern der Unigene-Datenbank zugeordnete SAGE-*tags*. In der Unigene-Datenbank sind nichtredundante Sequenz-Cluster von gut charakterisierten Transkripten und ESTs der Gen-Bank (GenBank, NCBI) enthalten, wobei jedes Cluster ein Transkript bzw. Gen repräsentiert. Die den Clustern zugeordneten SAGE-*tags* wurden durch Analyse aller Unigene-Sequenzen nach folgenden Prinzipien gewonnen (*Lash et al 2000*):

- Identifizierung eines PolyA-Signals (ATTAAG, AATAAG), eines PolyA-Sequenzteils (mindestens 8 A) und der 3' oder 5' Orientierung der Sequenz
- Extraktion eines 10 bp *tags*, welches 3' terminal an der letzten NlaIII-Restriktionsschnittstelle der Sequenz liegt
- Zuweisen des *tags* zum jeweiligen Unigene Cluster

Man unterscheidet den Modus „*full tag to gene mapping*“, in dem alle Unigene-Cluster unterschiedlicher Reliabilität in denen das jeweilige SAGE-*tag* gefunden wurde angezeigt werden, vom Modus „*reliable tag to gene mapping*“. In letzterem wurden 10 % der seltensten *tag*-Unigene Zuordnungen, die aufgrund einer Sequenzierfehler-Rate in den ESTs (*Hillier et al 1996*) als fehlerhaft angenommen werden, entfernt. Gefundene Sequenz-Cluster mit PolyA-Signal oder PolyA-Sequenzteil und 3'Orientierung besitzen die höchste Zuverlässigkeit, Sequenzen ohne PolyA-Signal oder -Sequenzteil und 3'Orientierung die geringste Zuverlässigkeit.

Bei der manuellen Recherche nach homologen EST-Sequenzen in den verfügbaren Sequenz-Datenbanken (z.B. GenBank, RIKEN, TIGR) unter Verwendung des BLAST-Algorithmus wurden für die Auswahl einer zuverlässigen homologen Sequenz die gleichen Bedingungen wie oben vorausgesetzt.

Alternativ zu einer Datenbankrecherche kann ein zu einem spezifischen *tag* gehörendes Transkript jedoch auch mit verschiedenen molekularbiologischen Techniken isoliert werden. Neben Hybridisierungstechniken, wie das von Velculescu beschriebene Colony Screening (*Velculescu et al 1995*), stellen PCR-basierte Methoden eine wichtige Möglichkeit zur

Identifikation der SAGE-*tag*-spezifischen Transkripte dar. Diese beruhen im Wesentlichen darauf, dass ein kurzer, *tag*-spezifischer 13-15 bp Vorwärtsprimer, bestehend aus der CATG(NlaIII)- und SAGE-*tag*-Sequenz, zusammen mit einem 3`terminalen Oligo(dT)- oder M13-Minusprimer an einer z.T. modifizierten, meist unverdauten Oligo(dT)cDNA eingesetzt wird (Polyak et al 1997, Peters et al 1999, Chen et al 2000). Van den Berg (Van den Berg et al 1999) favorisierte dabei eine Verlängerung des Vorwärtsprimers um fünf 5`Inosinreste zur Erhöhung der Annealingtemperatur. Bei Matsumura wurde die cDNA-Matrize in einen Vektor kloniert und die Amplifikation mit einem 13 bp *tag*-spezifischen Plusprimer und einer Vektorsequenz als Minusprimer durchgeführt (Matsumura et al 1999). In einem anderen modifizierten SAGE-Protokoll konnten längere 14 bp *monotags* generiert werden, welche sich gut als Primer für eine RACE-PCR eignen sollen (Ryo et al 2000). Ausschließlich zur Validierung der *tag*-Transkript-Identifikation ist eine *linker-anchored* PCR-Methode geeignet, bei der eine Sau3AI-restriktionsverdaute cDNA mit einem 5`terminalen Linker (sog. SADE-Linker) als Matrize dient und die Amplifikation mit einem 3` genspezifischen Minusprimer und einem 5` SADE-Linker spezifischen Plusprimer erfolgt (Cheval et al 2002, Piquemal et al 2002).

1.5.2. Aspekte der SAGE *tag*-spezifischen Transkriptidentifikation

In der Praxis zeigt sich bei der Analyse der SAGE-Daten immer wieder ein relevantes Problem: obwohl jedes *tag* theoretisch ausreichend Information für die korrekte Zuordnung zum korrespondierenden Gen enthält, ist die Identifizierung des zugrundeliegenden Transkriptes oft schwierig oder sogar unmöglich.

Die Homologierecherche in den verfügbaren EST-Datenbanken kann erfolglos verlaufen oder multiple Zuordnungen ergeben. Eine direkte Isolierung *tag*-korrespondierender Transkriptsequenzen mit den bisher etablierten molekularbiologischen Standardtechniken, welche auf modifizierten PCR-Protokollen oder Colony Screening mit kurzen Oligonukleotid-Proben basieren, ist aufgrund der kurzen *tag*-Sequenz (9-11 bp) schwer durchführbar, auch wenn auf diese Weise korrekte Klone erfolgreich identifiziert werden konnten (Velculescu et al 1995). Leider ist aber keine dieser Methoden zur Genidentifizierung in großem Umfang geeignet:

Das Screening mit kurzen Oligonukleotiden ist zeit- und arbeitsaufwendig. Die publizierten PCR-Methoden zur Genidentifizierung schlagen manchmal fehl oder müssen unter anspruchsvollen Amplifikationsbedingungen, die für jedes *tag* einer Optimierung bedürfen, durchgeführt werden.

Nicht nur wegen der erheblichen Kosten die zur Generierung von SAGE-Expressionsdaten notwendig sind - sondern auch, um die Möglichkeit der Entdeckung neuer Gene durch SAGE voll auszuschöpfen - ist jedoch eine zuverlässige und schnelle Methode zur Identifizierung der Kandidatengene nötig. Aus der Transkriptidentifizierung ergeben sich weitreichende Konsequenzen für spätere funktionelle Studien an den detektierten Genen.

Deshalb war die Etablierung einer robusten, in großem Umfang gleichermaßen effizient für alle *tag*-Sequenzen anwendbaren, molekularbiologischen Methode zur direkten Isolierung *tag*-korrespondierender Transkriptsequenzen erforderlich. Ziel ist, die Kosten die durch falsch positive oder falsch negative *tag*-zu-Gen Zuordnung verursacht werden, wesentlich zu minimieren. Wenn möglich, sollte eine solche Methode außerdem fähig sein, SAGE-Expressionsdaten zu bestätigen oder zu widerlegen.

1.5.3. Validierung von SAGE-Daten

SAGE stellt, wie alle anderen Screeningmethoden, nur den Ausgangspunkt eines komplexen methodischen Ansatzes zum Studium der Genexpression dar. Die Weiterverarbeitung der gewonnenen Daten beinhaltet deshalb eine zusätzliche unabhängige Analyse zur Bestätigung und zur Beseitigung falsch positiver Daten. Im Allgemeinen werden dafür Techniken wie der Northern Blot, der RNase Protection Assay, die in situ Hybridisierung und die (semiquantitative) RT-PCR angewendet.