

Molekulares Wechselspiel von Wirt und Pathogen: Simultane, genomweite Transkriptprofilierung zweier Organismen mit SuperSAGE

Günter Kahl¹, Peter Winter¹, Hideo Matsumura², Monika Reuter³, Detlev Krüger³ und Ryohei Terauchi²

¹Biozentrum der Universität Frankfurt am Main und GenXPro im Frankfurt Innovationszentrum (FIZ) Biotechnologie, Frankfurt am Main, ²Iwate Biotechnology Research Center, Kitakami, Japan, ³Universitätsklinikum Charité, Humboldt-Universität Berlin

► Obwohl die Totalsequenzierung der Genome einer zunehmenden Anzahl pro- und eukaryoter Organismen eine bislang nicht bekannte Fülle neuer Erkenntnisse mit sich bringt, lässt sich doch aus einer bestimmten Nukleotidsequenz nicht zwingend deren Funktion ableiten. So kann nur relativ wenigen Genen, die *in silico* identifiziert wurden, eine Funktion zugeordnet werden. Dies leistet erst die funktionelle Genomik, also die funktionelle Charakterisierung von Genen. Ein erster Schritt dazu ist die Feststellung, in welchen Geweben und unter welchen Bedingungen ein Gen in welcher Stärke exprimiert wird, das heißt, wo, wann und wie viele Transkripte davon abgelesen werden. Nachfolgend werden dann i. d. R. gentechnische Ansätze verwendet, bei denen die Gene überexprimiert oder stillgelegt werden, um aus den resultierenden Veränderungen des Phänotyps mehr über die Genfunktion zu erfahren.

Von hohem Interesse ist insbesondere die genetische Reaktion eines Organismus auf innere oder äußere Reize wie etwa Stress, Befall durch einen Schädling oder Krankheitserreger, aber auch die Reaktion auf Medikamente. Da solche Umweltveränderungen meist eine ganze Kaskade von Genen aktivieren oder inaktivieren, sind für das Verständnis eines bestimmten Phänotyps Techniken gefordert, die idealerweise die Expressionsmuster aller Gene messen. Diese genomweite Transkriptprofilierung liefert dann ein dynamisches Bild der Genaktivität während eines biologischen Prozesses.

Geschlossene und offene Transkriptomanalysen

Prinzipiell lassen sich solche Experimente mit geschlossenen und offenen Technologiesystemen durchführen. Geschlossene Systeme („closed architecture systems“) erlauben nur die Aktivität schon bekannter Gene festzustellen. Zu den meistgenutzten geschlossenen Systemen gehört die Micro-

array-Technik in ihren zahllosen Varianten. Solche Expressionsarrays sind hervorragend für die Hochdurchsatzanalyse der Transkription einer großen Anzahl von Genen unter verschiedenen experimentellen Bedingungen geeignet, erfassen dafür aber generell nur bekannte Gene. Dagegen sind offene Systeme („open architecture systems“) prinzipiell in der Lage, die Aktivität aller Gene einer Zelle zu messen, und – in Abhängigkeit von der verwendeten Technik – auch zu quantifizieren. Lediglich solche quantitativen Techniken entsprechen zukünftigen Qualitätsstandards. Hierzu gehören neben einer Reihe selten benutzter Methoden (z.B. RAGE^[1]; Übersicht in^[2]) vor allem zwei komplexe Techniken, nämlich „massively parallel signature sequencing“ (MPSS^[3]) und „serial analysis of gene expression“ (SAGE^[4]). Hier soll nur SAGE kurz porträtiert werden. SAGE produziert in mehreren Schritten von einer definierten Position am 3'-Ende eines jeden Transkripts (besser: der entsprechenden cDNA) 14 Basenpaare lange „tags“, von denen jedes „tag“ charakteristisch für die cDNA ist, von der es abgeleitet wurde. Die Anzahl verschiedener „tags“ reflektiert dabei die Anzahl der aktiven Gene, wogegen die relative Häufigkeit, mit der ein bestimmtes „tag“ erscheint, die Aktivität eines bestimmten Gens quantifiziert. Diese „tags“ werden letztlich gereinigt, zu möglichst langen Concatemeren aneinander ligiert, in geeignete Vektoren kloniert und sequenziert, und so die Anzahl der einzelnen „tags“ in einer Probe festgestellt. Es werden also alle Transkripte einer Zelle und deren Anzahl erfasst („quantitative global gene expression profiling“). Der Vorteil dieser Methode gegenüber der direkten Sequenzierung vieler, die Transkripte repräsentierenden cDNAs liegt darin, dass mit einer einzigen Sequenzierung 30 bis 40 cDNAs charakterisiert werden können, was bei gleicher Zuverlässigkeit die Kosten erheblich reduziert. Leider ist oft die geringe Länge eines „tags“ für die zweifelsfreie



Identifizierung eines Transkripts (oder seines Gens) unzureichend, weil ein „tag“ verschiedene Gene identifizieren und die Analyse verwirren kann. Daher wurde mit der sog. LongSAGE-Technik^[5] eine Verbesserung versucht. LongSAGE produziert zwar 19-21 Basenpaare lange „tags“, wird aber heute wegen methodischer Unzulänglichkeiten nur noch zur Annotation der SAGE-„tags“ benutzt.

SuperSAGE: Lange „tags“ für die verlässliche Transkriptom-Analyse

SAGE und LongSAGE können im Prinzip nur bei relativ wenigen Modellorganismen, für die große Mengen von „expressed sequence tags“ (ESTs) oder genomischen DNA-Sequenzen in den Datenbanken deponiert sind (also z.B. Hefe, Maus oder Mensch) sinnvoll angewendet werden. Für

Organismen ohne substantielle Sequenzinformation erfordert die Annotation von SAGE-„tags“ die Einbeziehung der cDNA-Sequenzen am 3'-Ende, wozu sich die sog. 3'-RACE-Methode (3' rapid amplification of cDNA ends) anbietet. Die 14Bp (SAGE) oder 19-21 Bp (LongSAGE) „tags“ sind aber zu kurz, um bei dieser Technik als spezifische PCR-Primer zu fungieren. Daher haben wir eine substantiell verbesserte Methode entwickelt, die auf der Verwendung des Typ III- Restriktionsenzym *EcoP15I* beruht, das 26 Bp entfernt von seiner Erkennungssequenz zum 3'-Ende der cDNA hin schneidet und damit ausreichend lange „tags“ produziert. Diese Verbesserung, verbunden mit einer drastischen Vereinfachung des konventionellen SAGE-Protokolls, erlaubt die schnelle und eindeutige Identifizierung der entsprechenden Gene und damit eine exakte Genexpressionsanalyse. Wir bezeichnen diese Technik als SuperSAGE^[6], die äußerst erfolgreich für eine genomweite quantitative Transkriptprofilierung eingesetzt wurde.

SuperSAGE: Gleichzeitige Analyse des gesamten Transkriptoms zweier Organismen

Im Folgenden soll das Potential von SuperSAGE am Beispiel der Reis pflanze und ihres Pilzpathogens *Magnaporthe grisea* skizziert werden. Dieser Ascomycet verursacht Läsionen an Reisblättern und ist für jährliche Schäden in Milliardenhöhe verantwortlich. In den Läsionen wächst das Hyphengeflecht des Pilzes in lebendes Blattgewebe ein, weshalb eine Trennung von Pathogen und Wirt unter natürlichen Bedingungen praktisch unmöglich ist. Dennoch konnte die Transkriptionsaktivität der Wirtspflanze und des Pathogens mit SuperSAGE zum ersten Mal eindeutig und quantitativ bestimmt werden. Da die Genomsequenz beider Organismen vollständig vorliegt, ist eine Zuordnung der „tags“ zu bestimmten Genen überdies sehr leicht. Das relativ einfache Protokoll beginnt mit der Isolierung von Gesamt-RNA, wovon die polyadenylierten Messenger-RNAs beider Organismen abgetrennt werden. Diese werden mit reverser Transkriptase und einem biotinylierten Oligo(dT)-Primer, der die Erkennungssequenz 5'-CAGCAG-3' von *EcoP15I* enthält in cDNA konvertiert, die dann mit der Restriktionsendonuklease *NlaIII* spezifisch geschnitten wird. Die 3'-Enden der cDNA-Fragmente werden an Streptavidin-haltige Magnetkügelchen gebunden und in zwei Portionen geteilt. In getrennten Ansätzen werden nun zwei fluoreszente Oligonukleotid-Linker (1E und 2E), beide mit der Erkennungssequenz von *EcoP15I*, mit T4-DNA-Ligase an die cDNA-Enden ligiert. Nach diesem Schritt wird al-

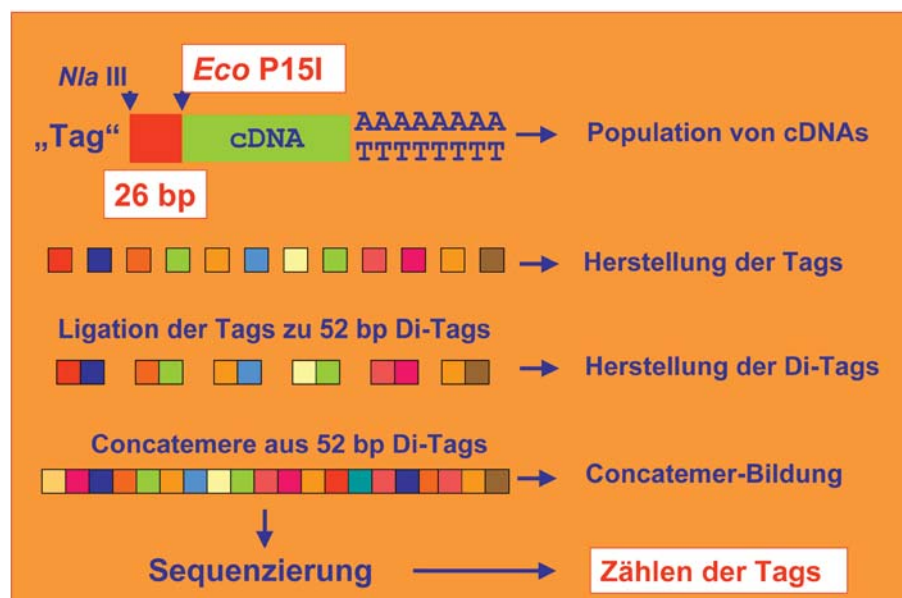


Abb. 1: Vereinfachtes Schema der SuperSAGE-Technik zur genomweiten quantitativen Transkriptanalyse.

Tab. 1: SuperSAGE: Quantitative Transkript-Profilierung von Wirt und Pathogen

Einige stark exprimierte Gene in Blattgewebe von Reis (<i>Oryza sativa</i> L.)		
„tag“-Sequenz	„tag“-Anzahl	Kodierte Proteine
CATGTTCCGGCTTCTTCGTCAGGCCA	122	D00641 : Chlorophyll-a/b-Bindeprotein
CATGGATCCGCTCTCTGGGAGGAAT	116	AU172529: Thiazol-synthetisierendes Enzym
CATGGCGACGCATCGCCTTCAGCTAA	114	X13909 : Chlorophyll-a/b-Bindeprotein
CATGTGGTGGCTTAGCTCTACGTGTA	111	AU174449: Glyzin-reiches Protein
CATGTCGGACAAGTGC GGCAACTGCG	94	TC56630 : Metallothionein
CATGTTGTAATACTCCATCAAAGAGT	86	D29966 : Katalase
CATGAATTGAGTTTCGCTTTGGTTATG	78	AF010579: Glyzin-reiches Protein
CATGATGATGATATACTACACTTGAT	58	BE230408: Photosystem II 10 Kd Protein
CATGGCGTCCACGCTGACCAACGTCG	57	BE230423: Unbekanntes Protein
CATGTATGTATGTACCTTAATTGTGT	52	D00642 : Chlorophyll-a/b-Bindeprotein
Gesamtanzahl der „tags“		12.057

Einige Transkripte von stark exprimierten Pilz-Genen in infiziertem Blattgewebe von Reis

„tag“-Sequenz	„tag“-Anzahl	Kodierte Proteine
CATGCGATCACGAGGGGATGATGGTG	38	Hydrophobin
CATGTCAGACACAGGCTGTACAAGGC	2	Nucleosid-Diphosphatkinase
CATGTCACGTTTAGAAAGCGACCCG	2	60S ribosomales Protein
CATGTTGCCCGTATGTACATAAACA	1	NADH-Ubichinon Oxidoreduktase
CATGCAATTGGTGTTCCTTTGGGTTT	1	Poly-Ubiquitin
CATGTCGCTGTGGCTTCAGTTGCTG	1	Unbekanntes Protein
⋮		⋮

