



(10) **DE 10 2008 025 656 B4** 2016.07.28

(12)

Patentschrift

(21) Aktenzeichen: **10 2008 025 656.0**

(22) Anmeldetag: **28.05.2008**

(43) Offenlegungstag: **03.12.2009**

(45) Veröffentlichungstag
der Patenterteilung: **28.07.2016**

(51) Int Cl.: **C12Q 1/68 (2006.01)**

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(73) Patentinhaber:
GenXPro GmbH, 60438 Frankfurt, DE

(74) Vertreter:
**Ackermann, Joachim Karl Wilhelm, Dipl.-Chem.
Dr.rer.nat., 60327 Frankfurt am Main, DE**

(72) Erfinder:
Rotter, Björn, Dr., 60320 Frankfurt, DE

(56) Ermittelter Stand der Technik:
siehe Folgeseiten

(54) Bezeichnung: **Verfahren zur quantitativen Analyse von Nukleinsäuren, Marker dafür und deren Verwendung**

(57) Hauptanspruch: Verfahren zur quantitativen Bestimmung von Nukleinsäuren in einer Probe dadurch gekennzeichnet, dass

a) Nukleinsäurekonstrukte, enthaltend

– mindestens eine kennzeichnende Polynukleotiddomäne einer, in einer Probe vorhandenen Nukleinsäure, die quantitativ bestimmt werden soll und

– mindestens eine artifizielle Polynukleotiddomäne mit zufälliger Sequenz und

– mindestens eine Primer- oder Polymerasebindungsstelle, wobei jeweils eine kennzeichnende Polynukleotiddomäne und mindestens eine Polynukleotiddomäne mit zufälliger Sequenz eine charakteristische Domänenkombination bilden die von mindestens einer Primer- oder Polymerasebindungsstelle flankiert wird,

in Anwesenheit einer Polymerase amplifiziert werden und
b) nach der Amplifikation zumindest die Sequenz der Polynukleotiddomänen mit zufälliger Sequenz und eines Teils der kennzeichnenden Polynukleotiddomäne ermittelt werden, wobei zur quantitativen Bestimmung der in der Probe vorhandenen Nukleinsäure sequenzidentische oder sequenzähnliche Kombinationen einer kennzeichnenden Polynukleotiddomäne und einer Polynukleotiddomäne mit zufälliger Sequenz als Kopien einer kennzeichnenden Polynukleotiddomäne der ursprünglich in der Probe vorhandenen Nukleinsäure identifiziert werden können.



(56) Ermittelter Stand der Technik:

US 2007 / 0 172 854 A1
US 2007 / 0 172 873 A1
US 2008 / 0 108 804 A1
WO 2004/ 015 085 A2
WO 2005/ 111 242 A2
WO 2006/ 003 721 A1

GOWDA, M. [u.a.]: Robust analysis of 5'-transcript ends (5'-Rate): a novel technique for transcriptome analysis and genome annotation. Nucleic Acids Res. (2006) 34 (19), e126, 1-9

MATSUMURA, H. [u.a.]: SuperSAGE. Cellular Microbiology (2005) 7 (11) 11-18

MEYER, M. [u.a.]: Parallel tagged sequencing on the 454 platform. Nat. Protoc. (2008) 3 (2) 267-78 (elektronisch veröffentlicht am 31.01.2008)

WONG, K.K. [u.a.]: Use of tagged random hexamer amplification (TRHA) to clone and sequence minute quantities of DNA-application to a 18 kb plasmid isolated from Sphingomonas F199. Nucleic Acids Res. (1996) 24 (19) 3778-3783

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur quantitativen Bestimmung von Nukleinsäuren in einer Probe, insbesondere zur quantitativen Bestimmung von Gentranskripten, genomischer DNA oder genomischen Nukleinsäureabschnitten, wie z. B. DNA, RNA, mRNA, cDNA, microRNA, non-codingRNA, Aptamer-DNA und -RNA, sowie die Bereitstellung von Markern zur Durchführung der Analyseverfahren.

[0002] Die Quantifizierung von Nukleinsäuren hat große wissenschaftliche und kommerzielle Bedeutung in allen Bereichen der Lebenswissenschaften. So ist zum einen wichtig zu wissen, welche Genprodukte (z. B. für Proteine kodierende und nicht-kodierende RNA) wie häufig vorkommen (Genexpression-Studien), zum anderen wie die Transkription reguliert wird, z. B. durch epigenetische Eigenschaften der genomischen DNA. Zudem ist die Bestimmung der Häufigkeit von Kopien von Genabschnitten zur Karyotypisierung eine wichtige Diagnosemöglichkeit von Krankheiten wie z. B. Krebs.

[0003] Dieser Bedarf hat zur Entwicklung verschiedener Verfahren geführt, die in der Regel entweder auf der unspezifischen Messung der Gesamtmenge aller oder bestimmter Nukleinsäuren durch z. B. photometrische Methoden (Gesamt-RNA oder DNA) beruhen, oder es sollen eine Vielzahl verschiedener Nukleinsäure-Moleküle bestimmter Sequenz individuell quantifiziert werden. Die dazu verwendeten Methoden können grundsätzlich in 2 Gruppen eingeteilt werden: Die erste Gruppe bilden Verfahren, die schon bekannte oder anderweitig charakterisierte Sequenzen quantifizieren können. Sie basieren in der Regel auf der Bindung der zu quantifizierenden Nukleinsäuren an bekannte oder anderweitig charakterisierte Sequenzen, die an eine feste Oberfläche gebunden sind. Beispiele sind alle Formen des Southern-Blots, Northern-Blots oder sogenannte DNA- oder RNA-Micro-Arrays (Chips).

[0004] Die zweite Gruppe bilden Verfahren, die auf der Quantifizierung der Nukleinsäuren durch Sequenzierung möglichst vieler individueller Nukleinsäure-Moleküle und Auszählen der sequenzierten Moleküle im Gemisch beruhen. Da in der Regel mehr Nukleinsäuremoleküle im Gemisch vorhanden sind, als sequenziert werden, wird so die relative Häufigkeit einer Nukleinsäure bestimmter Sequenz im Gemisch erhalten.

[0005] Die hier vorgestellte Erfindung verbessert die Sicherheit und Genauigkeit der Quantifizierung von Nukleinsäuren durch nahezu alle auf Sequenzierung und Auszählung basierenden Verfahren. Darunter fallen sowohl so genannte „Tag“-basierte Verfahren, wobei der Tag ein für die Nukleinsäure möglichst repräsentatives Teilstück darstellt, anhand dessen die Nukleinsäure identifiziert werden kann, als auch Verfahren, bei der die Anzahl kompletter DNA- und RNA-Moleküle wie etwa die von Viren bestimmt wird.

[0006] Sequenz-basierte, quantitative Verfahren finden aufgrund der Entwicklung von Techniken zur gleichzeitigen Sequenzierung von Hunderttausenden bis Millionen von Nukleinsäure-Molekülen, immer breitere Anwendung und ersetzen qualitative, Array-basierten Verfahren in vielen Bereichen. Deshalb wird die Sequenz-basierte Quantifizierung in Zukunft einen immer höheren Stellenwert erhalten. Eine Liste der wichtigsten Sequenz-basierten Quantifizierungsmethoden für Nukleinsäuren findet sich in S. M. Wang (2007) Understanding SAGE data, Trends in Genetics 23: 42–50. Im Einzelnen gehören dazu:

- „Serial analysis of gene expression“ (SAGE, V. E. Velculescu et al., Serial analysis of gene expression, Science 270 (1995), pp. 484–487, US-Patent 5,695,937): SAGE erleichtert die globale, quantitative Charakterisierung eines Transcriptoms indem es ein Typ-II Restriktionsenzym (BsmFI) als „Tagging Enzym“ verwendet, um aus einer cDNA ein 14-Bp langes Fragment, den sogenannten „Tag“ hinter der am weitesten zum 3'-Ende gelegenen Schnittstelle für ein häufigschneidendes Restriktionsenzym (meist hinter der NlaIII-Schnittstelle CATG) herauszuschneiden. Bei SAGE werden jeweils 2 Tags Kopf-an Kopf zu so genannten Ditags zusammenligiert, die dann zu längeren Ketten („Konkatemere“) verbunden („konkatemerisiert“) werden. Diese Konkatemere werden dann kloniert und sequenziert. Die Tags repräsentieren die AusgangsmRNA-Moleküle, von denen die cDNA generiert wurde. Daher stellt die Menge der sequenzierten Tags ein Maß für die relative Häufigkeit dar, mit der die mRNA im mRNA-Gemisch vertreten war. Heute kann man auf Konkatemer-Bildung und Klonierung verzichten, weil die Ditags nach einer PCR direkt mit hochparallelen Sequenzierungstechniken wie etwa den oben beschriebenen Verfahren sequenziert werden können (z. B. Nielsen KL, Høgh AL, Emmersen J., Nucleic Acids Res. 2006; 34(19): e133. DeepSAGE-digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples).
- „LongSAGE“ (S. Saha et al., Using the transcriptome to annotate the genome, Nat. Biotechnol. 20 (2002), pp. 508–512) wurde entwickelt um die Spezifität der nur 14 Bp langen SAGE-tags zu erhöhen, indem hier durch Verwendung eines anderen Typ-II-Restriktionsenzym (MmeI) 21 Bp lange tags generiert werden. Dadurch wird die Zuordnung zu genomischen oder EST-Sequenzen verbessert. Die Quantifizierung mit dem SAGE-Verfahren und insbesondere mit der LongSAGE ist problematisch, so ist es z. B. oft nö-

tig vermehrungsbedingte Verfälschungen der Quantifizierung durch biostatistische Analyseverfahren (z. B. Emmersen, J. et al., BMC Bioinformatics 2007, 8: 92) zu beheben. Diese Verfahren beruhen jedoch auf statistischen Annahmen und können daher die tatsächliche relative Häufigkeit der Tags in einer Probe nur ungefähr abschätzen.

- "SuperSAGE" (H. Matsumura et al., Gene expression analysis of plant host-pathogen interactions by SuperSAGE, Proc. Natl. Acad. Sci. U.S.A. 100 (2003), pp. 15718–15723) bringt durch Verwendung des Typ-III-Restriktionsenzyms EcoP15I als Tagging-Enzyme 25–27 Bp lange Tags hervor, die neben der weiter verbesserten Zuordnung zu anderen Sequenzen weitere Vorteile besitzen.

- "Cap analysis gene expression" (CAGE, T. Shiraki et al., Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage, Proc. Natl. Acad. Sci. U.S.A. 100 (2003), pp. 15776–15781) wurde entwickelt, um Transcript-Initiationssequenzen und Promotoren zu identifizieren. Bei dem Verfahren werden 21 Bp lange Fragmente isoliert und sequenziert, die direkt an die 5'-CAP-Sequenz einer mRNA anschließen.

- "Gene identification signature" (GIS, C. L. Wei et al., 5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation, Proc. Natl. Acad. Sci. U.S.A. 101 (2004), pp. 11701–11706) dient dazu, gleichzeitig die 5'- und 3'-Enden von mRNAs zu bestimmen. GIS-Tags sind 20 Bp lang und können dazu verwendet werden, die Enden eines Gens im Genom zu identifizieren.

- „Digitale Karyotypisierung" DK; (T. L. Wang et al., Digital karyotyping, Proc. Natl. Acad. Sci. U.S.A. 99 (2002), pp. 16156–16161) adaptiert das LongSAGE-Protokoll um 21 Bp-tags aus genomischer DNA zu gewinnen. Damit lässt sich die krankhafte Vermehrung oder der Verlust bestimmter Chromosomenabschnitte sowie Insertionen und Fremd-DNA z. B. in Krebszellen genau bestimmen (J. T. Park et al., Notch3 gene amplification in ovarian cancer, Cancer Res. 66 (2006), pp. 6312–6318).

- „Methylierungsspezifisches Digitales Karyotypisieren" (MSDK) (Min Hu et al., Methylation-specific digital karyotyping Nature Protocols 1, – 1621–1636 (2006)) wendet das Prinzip des Digital Karyotyping für die quantitative Bestimmung unterschiedlicher Methylierungszustände zweier Proben an.

- "Paired-end ditag" (C. L. Wei et al., A global map of p53 transcription-factor binding sites in the human genome, Cell. 124 (2006), pp. 207–219) zielt darauf ab, Protein-bindende Sequenzen im Genom zu identifizieren. Dabei wird das Prinzip von GIS dazu verwendet, Tags von beiden Enden eines DNA-Fragments zu isolieren und zu quantifizieren, die an ein bestimmtes Protein (z. B. einen Transkriptionsfaktor) gebunden sind, das zusammen mit der gebundenen DNA durch Immuno-Präzipitation spezifisch aus dem Protein-DNA-Gemisch gefällt wird. Paired-end ditag wurde verwendet, um p53-Bindestellen im Menschengenom zu identifizieren.

- Metagenomische Analysen verfolgen das Ziel, die Zusammensetzung einer komplexen Mischung von Organismen zu ermitteln. Dabei werden dieselben Verfahren wie bei der Digitalen Karyotypisierung verwendet (Krause et al., Phylogenetic classification of short environmental DNA fragments, Nucleic Acids Res. 2008 April; 36(7): 2230–2239).

[0007] Auch in US 2007/0172973 wird ein Verfahren zur quantitativen Bestimmung von Nukleinsäuren beschrieben, wobei die zu quantifizierenden Nukleinsäuren anhand ihrer Markierung erst sortiert und dann deren Mengen bestimmt werden.

[0008] In allen beschriebenen Verfahren reicht die zur Verfügung stehende Menge an Nukleinsäuren einer Probe meist nicht für alle nachfolgenden Analysen aus. Darum ist vor der Analyse fast immer eine Vermehrung des Materials notwendig.

[0009] Beispielsweise benötigen hochparallele Sequenzierverfahren (z. B. das „454" Picoliter Verfahren von Roche Diagnostics, Deutschland, oder das Solexa-Pyrosequenzierverfahren der Firma Illumina, Inc., San Diego, USA) 5–3000 ng DNA. Zudem benötigen beide Verfahren spezifische Sequenzen (Adapter) an beiden Enden der zu sequenzierenden Nukleinsäure, die als Primerbindungsstelle dienen. Um das Vorhandensein dieser spezifischen Sequenzen an allen zu sequenzierenden Molekülen zu gewährleisten, wird die Polymerase-Kettenreaktion (PCR) benutzt, wobei Oligonukleotide die den Adaptersequenzen komplementär sind, als Primer Verwendung finden. Auch für eine Klonierung der Fragmente ist oft ihre vorhergehende Amplifikation notwendig.

[0010] Die Amplifikation, z. B. mittels PCR, birgt jedoch die Gefahr, dass bestimmte Nukleinsäuren präferentiell vermehrt werden. So werden kürzere Nukleinsäuren bevorzugt, aber auch die individuelle Basenabfolge kann die Effizienz der Amplifikation beeinflussen. Dadurch kann die relative Häufigkeit einer Nukleinsäure in einer Nukleinsäurepopulation vor und nach der Amplifikation stark voneinander abweichen.

[0011] Die Quantifizierung von Nukleinsäure-Zusammensetzungen wird zudem häufig, wie etwa bei der Quantifizierung von Expressed-Sequenced Tags (ESTs), auch ohne vorherige PCR-Amplifikation z. B. nach einer Klonierung der Nukleinsäuren durchgeführt. Auch hierbei kann aufgrund unterschiedlicher Fitness individueller Bakterien und aufgrund von Effekten der Nukleinsäuren auf das Wachstum der transformierten Zellen eine Verfälschung der ursprünglichen Zusammensetzung der Nukleinsäurepopulation entstehen.

[0012] Aufgabe der vorliegenden Erfindung war es nun ein einfaches Verfahren zur Verfügung zu stellen, bei dem die während der Vermehrung der zu quantifizierenden Nukleinsäuren in einer Probe auftretenden Verfälschungen der Nukleinsäure-Zusammensetzungen besser erkannt und bei einer quantitativen Auswertung korrigiert werden können.

[0013] Die Aufgabe wird dadurch gelöst, dass alle Nukleinsäuren eines Nukleinsäuregemischs, egal ob doppelsträngige oder einzelsträngige, vor der Amplifikation mit einem „Random Tag“ (RAT) versehen werden, der aus einem Gemisch aus synthetischen Oligonukleotiden stammt, wobei das Gemisch genügend RATs mit unterschiedlichen Sequenzen enthält um zu gewährleisten, dass die entstehenden Random-Tag-Nukleinsäuren (RAT-NS)-Kombinationen praktisch einzigartig sind. Die Methoden, wie die Nukleinsäuren vor der Amplifikation mit einem RAT versehen werden können, können sehr unterschiedlich sein. Nach einer Sequenzierung können die RAT-Nukleinsäure-Kombinationen ermittelt werden, es entsteht ein Datensatz mit den individuellen Sequenzinformationen der RAT-NS. RAT-Nukleinsäure-Kombinationen, die mehr als einmal oder mindestens häufiger als statistisch wahrscheinlich vorkommen sind somit als Kopien erkennbar. Eliminiert man alle Kopien aus dem Datensatz, so ergibt sich die ursprüngliche Zusammensetzung des Nukleinsäuregemischs vor der Amplifikation.

[0014] Folglich ist der Gegenstand der vorliegenden Erfindung ein Verfahren zur quantitativen Bestimmungen von Nukleinsäuren (Nukleinsäurepopulation) in einer Probe bei dem

a) Nukleinsäurekonstrukte, enthaltend

– mindestens eine kennzeichnende Polynukleotiddomäne (TAG) einer, in einer Probe vorhandenen Nukleinsäure, die quantitativ bestimmt werden soll und

– mindestens eine artifizielle Polynukleotiddomäne mit zufälliger Sequenz („Random-Tag“ oder abgekürzt RAT) und

– mindestens einer Primer- oder Polymerasebindungsstelle (PB)

wobei jeweils eine kennzeichnende Polynukleotiddomäne (TAG) und mindestens eine Polynukleotiddomäne mit zufälliger Sequenz (RAT) eine charakteristische Domänenkombination (TAG-RAT-Kombination) bilden, die von mindestens einer Polymerase- oder Primerbindungsstelle (PB) flankiert wird, in Anwesenheit einer Polymerase amplifiziert werden und

b) nach der Amplifikation zumindest die Sequenz der Polynukleotiddomänen mit zufälliger Sequenz (RAT) und zumindest eines Teils der kennzeichnenden Polynukleotiddomäne (TAG) ermittelt werden, wobei zur quantitativen Bestimmung der in der Probe vorhandenen Nukleinsäure sequenzidentische und gegebenenfalls zusätzlich sequenzähnliche Kombinationen einer kennzeichnenden Polynukleotiddomäne (TAG) und einer Polynukleotiddomäne mit zufälliger Sequenz (RAT) als Kopien einer kennzeichnenden Polynukleotiddomäne (TAG) der ursprünglich in der Probe vorhandenen Nukleinsäure identifiziert werden können.

[0015] Als „Nukleinsäure“ im Sinne der vorliegenden Erfindung werden die quantitativ zu bestimmenden Nukleinsäuren verstanden, die in einer Probe vorliegen. Die zu bestimmenden Nukleinsäuren können in der Probe sowohl einzelsträngig als auch doppelsträngig vorliegen. Dabei handelt es sich um DNAs und RNAs, insbesondere um mRNAs, microRNAs, rRNA und nicht kodierende RNAs sowie deren, durch reverse Transkription erzeugte cDNAs, die für eine Genexpressionsanalyse und eine Analyse der nicht-codierenden RNAs herangezogen werden können. Zudem umfasst der Begriff „Nukleinsäure“ auch Aptamer-DNA und -RNA bzw. Fragmente genomischer DNA, die insbesondere auch methyliert sein kann. Weiter zielt die Erfindung auf die Quantifizierung von durch Immunpräzipitation gewonnenen DNA Fragmenten und von Fragmenten, die durch Digital Karyotyping und Methylation-Specific Digital Karyotyping quantifiziert werden. Ebenso verbessert die Erfindung die Quantifizierung von DNA-Fragmenten für metagenomische Analysen sowie zur Analyse von Fragmenten aus subtraktiven DNA-Bibliotheken. Ebenfalls fallen unter den Begriff „Nukleinsäuren“ einzel- oder doppelsträngige Polynukleotide künstlichen (synthetischen) Ursprungs, darunter auch solche, deren Phosphat-Gruppen durch andere Atome ersetzt bzw. deren Basen modifiziert wurden und die mittels Polymerisation amplifizierbar und die sequenzierbar sind.

[0016] Als „Primer- oder Polymerasebindungsstelle (PB)“ wird eine Nukleinsäure bekannter Sequenz bezeichnet, die bevorzugt als Bindestelle für einen Primer z. B. für die PCR dienen kann, wobei die Amplifikation unter Verwendung der entsprechenden Primer und Polymerasen durch enzymatische Polymerisation erfolgt. Wei-

terhin können auch bekannte Sequenzen für eine Polymeraseanbindung zur Initiation einer enzymatischen Amplifikation benutzt werden, wie sie z. B. in bekannten Plasmiden oder anderen Vektoren bereits enthalten sind. In einem solchen Fall kann, unter Verwendung einer Insertionssequenz, der Einbau der RAT-TAG-Kombinationen in einen Vektor erfolgen, wodurch ein Nukleinsäurekonstrukt im Sinne der vorliegenden Erfindung entsteht, dass durch Klonierung vermehrt werden kann. Bei der Polymerasebindungsstelle kann es sich z. B. um eine Promotorsequenz für eine RNA-Polymerase, etwa um SP6 oder T7-RNA-Polymerase handeln, wobei von der Bindungsstelle aus eine lineare Amplifikation des Nukleinsäure-Konstrukts durch die entsprechende Polymerase erfolgen kann.

[0017] Unter einer „kennzeichnenden Polynukleotiddomäne“, wird ein sogenannter „TAG“ verstanden, nämlich eine Nukleinsäuredomäne deren Sequenz kennzeichnend für eine in der Probe vorkommende und quantitativ zu bestimmende Nukleinsäure ist. Welcher Teil einer Nukleinsäure als kennzeichnende Polynukleotiddomäne ausgewählt wird ist prinzipiell frei wählbar, solange die einzelnen TAGs einer bestimmten, in der Probe vorliegenden Nukleinsäure zugeordnet werden können. Ist diese Zuordnung nicht eindeutig, können weitere Verfahren angewendet werden, um diese Eindeutigkeit zu gewährleisten, z. B. mit Real time PCR mit spezifischen Primern. Als TAG können z. B. konservierte Teile einer Gensequenz genutzt werden, bevorzugt werden aber kennzeichnende Polynukleotiddomänen mit einer ausreichenden Sequenzlänge benutzt. Bevorzugt sollten die kennzeichnenden Polynukleotiddomänen eine Sequenzlänge von mindestens 8 Nukleotiden, bevorzugt von mindestens 12 Nukleotiden, besonders bevorzugt von mindestens 15 Nukleotiden besitzen. Eine kennzeichnende Polynukleotiddomäne kann dabei auch die gesamte zu bestimmende Nukleinsäure sein. Bevorzugt sollte die Domäne allerdings eine Länge von 50 Nukleotiden, bevorzugt von 35 Nukleotiden, insbesondere von 27 Nukleotiden nicht übersteigen. Dabei können auch längere, kennzeichnende Polynukleotiddomänen in ein Nukleinsäurekonstrukt eingebaut werden, wobei es allerdings häufig ausreicht, nur einen Teil dieser Sequenz im Anschluss an eine Amplifizierung der Konstrukte zu bestimmen.

[0018] Eine „Polynukleotiddomäne mit zufälliger Sequenz“ im Sinne der vorliegenden Erfindung ist ein synthetisch hergestellter (artificialer) „Random Tag“ (RAT), dessen Sequenz eine eindeutige Markierung der kennzeichnenden Polynukleotiddomäne zulässt. Die RATs werden üblicherweise kommerziell durch eine „statistische“ Verknüpfung der einzelnen Nukleotide hergestellt. Es wird dabei eine Mischung aus RATs erhalten, wobei die Wahrscheinlichkeit des Vorliegens von RATs mit gleicher Sequenz bei steigender Sequenzlänge und mit abnehmender Anzahl an RAT-Molekülen in der Mischung sinkt. Die so hergestellten Polynukleotiddomänen besitzen eine zufällige Sequenz im Sinne der Erfindung. Aufwendiger ist die Herstellung von RATs mit einer bekannten Sequenz, wobei so RATs hergestellt werden können, deren Sequenz garantiert einmalig ist. Auch solche RATs stellen eine „zufällige Sequenz“ im Sinne der vorliegenden Erfindung dar. Auch können RATs aus einer Mischung bekannter, unterschiedlicher Sequenzen hergestellt werden.

[0019] Einzelsträngige RATs können als Oligonukletide hergestellt oder erworben werden. Zur Herstellung von doppelsträngigen RATs werden bevorzugt einzelsträngige Oligonukleotide mit einer zufälligen Sequenz (RAT) und einem bekannten Sequenzanteil verwendet, wobei der bekannte Sequenzanteil am 3'-Ende des einzelsträngigen Oligonukleotides liegt. An den bekannten Sequenzanteil werden Oligonukleotide mit komplementärer Sequenz als Primer hybridisiert und mit Hilfe einer Polymerase (z. B. mit dem so genannten Klenow-Fragment der DNA-Polymerase I) in ein doppelsträngiges Oligonukleotid überführt, dass einen doppelsträngigen RAT enthält. Reine doppelstränge RATs ohne weitere Domänen können Integration einer Tagging-Enzym-Site und einer Primerbindungsstelle erzeugt werden, wobei nach Zweitstrangsynthese mit Hilfe des Tagging-Enzyms der RAT abgespalten wird. Weiterhin kann der einzelsträngige RAT bevorzugt an einem Ende mit einer bekannten Sequenz versehen werden, die bestimmte Polynukleotiddomänen wie eine Tagging-Enzym-Site, Restriktionssite, Ligationsstelle etc. enthalten kann. Am anderen Ende des RAT wird dann bevorzugt eine Primerbindungsstelle eingefügt, die auch zur Synthese des Doppelstranges mit dem Klenow Fragment herangezogen werden kann. Um RATs mit geeigneten Ligationsstellen (LS) zu versehen, die zu überhängenden Restriktionsschnittstellen komplementär sind, können Oligonukleotide die eine Primerbindungsstelle gefolgt von einem RAT und der entsprechenden Restriktionsschnittstelle gefolgt von einem weiteren RAT und einer weiteren Primerbindungsstelle enthalten, hergestellt werden. Es wird ein Konstrukt mit folgendem Aufbau erhalten: PB-RAT-Restriktionsschnittstelle-RAT-PB. Nach der Erzeugung des doppelsträngigen Konstruktes werden die Oligonukleotide mit dem entsprechenden Restriktionsenzym verdaut, es werden aus einem Oligonukleotid jeweils zwei RATs erhalten die mit überstehenden Enden für die Ligation modifiziert sind.

[0020] Außerdem können die RATs durch Hybridisierung zweier aus einer Zufallskombination bestehenden einzelsträngigen Oligonukleotiden hergestellt werden, die unter stringenten Bedingungen miteinander hybridisiert werden. Bei RATs die aus bekannten Sequenzen aufgebaut sind, ist dies die bevorzugte Methode.

[0021] Als „Nukleinsäurekonstrukte“ werden Polynukleotide verstanden, die zumindest eine kennzeichnende Polynukleotiddomäne, die für eine in der Probe enthaltene und quantitativ zu bestimmende Nukleinsäure kennzeichnend ist (TAG), eine Polynukleotiddomäne, mit einer zufälligen Sequenz (RAT) und mindestens einer Primer- oder Polymerase-Bindungsstelle (PB) enthalten. Darüber hinaus können weitere Polynukleotiddomänen in den Nukleinsäurekonstrukten enthalten sein, wie z. B. Linkersequenzen, Ligationsstellen (LS), Restriktionsenzymchnittstellen oder Tagging-Enzym-Bindungsstellen (TES). In einer bevorzugten Abfolge der Polynukleotiddomänen sind die kennzeichnende Polynukleotiddomäne (TAG) und die Polynukleotiddomäne mit zufälliger Sequenz (RAT) kovalent miteinander verknüpft. Die Verknüpfung kann durch direkte Ligation der einzelnen Domänen erfolgen. Alternativ können die einzelnen Domänen auch über Linkersequenzen miteinander verbunden werden, wobei sich die Linkersequenz durch die Ligation von korrespondierenden, überhängenden Enden ergibt. Die miteinander verknüpften TAG- und RAT-Domänen werden schließlich durch mindestens eine Primer- oder Polymerase-Bindungsstelle (PB) ergänzt, die bevorzugt kovalent, entweder an die kennzeichnende Polynukleotiddomäne (TAG) oder an die Polynukleotiddomäne mit zufälliger Sequenz (RAT) gebunden ist. In einer bevorzugten Ausführungsform liegen die TAG-Domäne und die RAT-Domäne flankiert von zwei verschiedenen Primerbindungsstellen (PB, PB') vor. Neben einer direkten Verknüpfung der Primer- oder Polymerasebindungsstellen mit den jeweiligen Enden der TAG-RAT-Domänenkombination kann auch eine Verknüpfung über einen Polynukleotidlinker in Frage kommen. Auch die Primer- oder Polymerasebindungsstellen (PBs) können durch Ligation in die Nukleinsäurekonstrukte eingebunden werden. Alternativ können TAGs aber insbesondere auch RATs eingesetzt werden, die schon im Vorfeld mit einer Primerbindungsstelle (PB) versehen sind.

[0022] In einer alternativen Ausführungsform können die zur enzymatischen Amplifikation vorbereiteten Nukleinsäurefragmente neben mindestens einer kennzeichnenden Polynukleotiddomäne (TAG) auch zwei oder mehr Polynukleotiddomänen mit zufälliger Sequenz (RATs) enthalten. Dabei können ein, zwei oder mehr Kombinationen aus einem TAG und einem RAT in einem Nukleinsäurekonstrukt enthalten sein. Die Orientierung der TAG-RAT-Kombinationen im Konstrukt spielt dabei keine Rolle. Weiterhin können einzelne TAG-RAT-Kombinationen unabhängig voneinander von einer Primer- oder Polymerase-Bindungsstelle (PB) flankiert vorliegen. Möglich ist allerdings auch, dass zwei Primer-Bindungsstellen (PBs) alle TAG-RAT-Kombinationen in einem Nukleinsäurekonstrukt einschließen bzw. eine Primer- oder Polymerase-Bindungsstelle (PB) an einem Ende der TAG-RAT-Kombinationen liegt.

[0023] So sind insbesondere Nukleinsäurekonstrukte mit folgender Polynukleotiddomänenabfolge von besonderem Interesse:

- a) PB-TAG-RAT
 - b) PB-RAT-TAG
 - c) PB-TAG-RAT-PB'
 - d) PB-RAT-TAG-RAT
 - e) PB-RAT-TAG-RAT-PB'
 - f) PB-RAT-TAG-TAG-RAT
 - g) PB-RAT-(TAG)_n-PB; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10
 - h) PB-RAT-(TAG)_n-RAT-PB; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10
 - i) PB-RAT-(TAG)_n; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10
 - j) PB-RAT-TAG-RAT-TAG
 - k) PB-TAG-RAT-TAG-RAT
 - l) PB-RAT-TAG-RAT-TAG-PB'
 - m) PB-(RAT-TAG)_n-PB'; mit n = 3 bis 50, bevorzugt mit n = 3 bis 10
 - n) PB-RAT-TAG-PB'-PB''-RAT-TAG-PB'''
 - so) PB-RAT-TAG-PB'-PB''-RAT-TAG-PB'''-PB''''-RAT-TAG-PB'''''
 - p) PB-RAT-TAG-PB'-(PB-RAT-TAG-PB')_n-PB-TAG-RAT-PB'; mit n = 0 bis 10
- (PB, PB', PB'', PB''', PB'''' = Primerbindungsstellen, wobei die Primerbindungsstellen bevorzugt unterschiedliche Sequenzen haben aber auch die gleiche Sequenz besitzen können.)

[0024] Dabei sind Nukleinsäurekonstrukte gemäß den Ausführungsformen a) b) c) e) und g) und h) besonders bevorzugt im Sinne der vorliegenden Erfindung.

[0025] Nukleinsäurekonstrukte können einzel- oder doppelsträngig vorliegen, wobei bei einzelsträngigen Nukleinsäurekonstrukten unter einer Primer-Bindungsstelle auch eine Sequenz verstanden wird, die einer Primer-Bindungsstelle auf dem Gegenstrang entspricht, während in einem doppelsträngigen Nukleinsäurekonstrukt sowohl die Primer-Bindungsstelle als auch deren korrespondierende Sequenz auf dem Gegenstrang enthalten ist.

[0026] Die Nukleinsäurekonstrukte dienen als Matrizen (Templates) für deren enzymatische Amplifikation durch eine Polymerase.

[0027] Die Nukleinsäurekonstrukte können auch weitere Oligonukleotiddomänen enthalten, wie z. B. Linker, Ligationsstellen, Restriktionssites und Tagging-Enzym Bindungsstellen (TES).

[0028] In einer weiteren bevorzugten Ausführungsform enthält das Nukleinsäurekonstrukt oder Vorstufen dieses Konstruktes eine oder mehrere Bindungsstellen für ein Tagging-Enzym (TE). Eine Tagging-Enzym-Site (TES) stellt eine Bindungsstelle für ein entfernt von seiner Erkennungstelle schneidendes Typ-II oder Typ-III Restriktionsenzym, bevorzugt für EcoP15I, Mmel oder Bsmfl, dar. Durch Einsatz eines Tagging-Enzyms können insbesondere Nukleinsäurefragmente mit einer definierten Länge erzeugt werden, die als kennzeichnende Nukleotid-Domäne (TAG) dient. Die Tagging-Enzym-Site (TES) ist so einzubauen, dass der benötigte TAG auch nach der enzymatischen Spaltung im Nukleinsäurekonstrukt bzw. in einer Konstruktvorstufe enthalten ist. Bevorzugt liegt die Bindungsstelle für ein Tagging-Enzym (TES) am distalen Ende zur PB-Sequenz des RAT, die Einbindung der TES in ein Nukleinsäurekonstrukt bzw. in eine entsprechende Konstruktvorstufe erfolgt somit bevorzugt in der Domänenabfolge „PB-RAT-TES-Nukleinsäure“. Das Produkt der späteren Spaltung des Nukleinsäurekonstrukts mit dem TE ist daher ein Polynukleotid mit der Domänen-Abfolge PB-RAT-TES-TAG. Alternativ kann aber auch zuerst der TAG mit Hilfe des Tagging-Enzyms erzeugt werden und danach wird der RAT an einem beliebigen Ende der TES-TAG-Domänenabfolge angebunden. Wird z. B. EcoP15I als Tagging-Enzym verwendet, ist die Anwesenheit zweier Kopf an Kopf orientierter Erkennungsstellen (Tagging-Enzym-Sites) mit der Sequenz CAGCAG nötig, welche die Schnittstelle des Tagging-Enzyms einschließen und bevorzugt in einem Abstand von weniger 3000 Bp liegen. Die bevorzugte Domänen-Abfolge ist daher im Falle der Verwendung eines Typ III Restriktionsenzyms, wie EcoP15I, „PB-RAT-TES-Nukleinsäure-TES“. Das Produkt der späteren Spaltung des Nukleinsäurekonstrukts mit EcoP15I als TE sind daher ein Polynukleotid mit der Domänen-Abfolge PB-RAT-TES-TAG, und ein Polynukleotid mit der Domänen-Abfolge TES-Tag, von denen das PB-RAT-TES-TAG-Konstrukt bevorzugt für die Charakterisierung der Nukleinsäure verwendetet wird. Ein großer Vorteil bei der Verwendung von EcoP15I ist, dass besonders lange TAGs mit 25–27 Basenpaaren erhalten werden.

[0029] In einer anderen, bevorzugten Anwendungsform werden beide Enden des zu charakterisierenden Nukleinsäuremoleküls mit gleichen oder verschiedenen PB-RAT-TES-Konstrukten versehen. Die Spaltung des Konstrukts mit dem TE resultiert dann in zwei Molekülen mit der Domänen-Abfolge PB-RAT-TES-TAG, wobei die PB und TES-Domänen gleiche oder unterschiedliche Sequenz haben können. In diesem Fall können zwei PB-RAT-TES-TAG-Konstrukte zur Charakterisierung des Nukleinsäuremoleküls verwendet werden.

[0030] Nach Schneiden mit dem entsprechenden Tagging-Enzym können dann weitere Polynukleotiddomänen, wie z. B. RATs und PBs, an den erhaltenen TAG angebunden werden, so dass sich in den resultierenden Nukleinsäurekonstrukten, die folgenden besonders bevorzugten Domänenabfolgen ergeben:

- a') PB-TES-TAG-RAT
 - a'') PB-TAG-TES-RAT
 - b') PB-RAT-TAG-TES
 - b'') PB-RAT-TES-TAG
 - c') PB-TAG-TES-RAT-PB'
 - c'') PB-TES-TAG-RAT-PB'
 - d') PB-RAT-TAG-TES-RAT
 - d'') PB-RAT-TES-TAG-RAT
 - e') PB-RAT-TAG-TES-RAT-PB'
 - e'') PB-RAT-TES-TAG-RAT-PB'
 - f') PB-RAT-(TES)_m-TAG-TAG-(TES)_m-RAT;
- mit m = 0 oder 1 unabhängig voneinander für jedes m, wobei mindestens ein m = 1 ist
- g') PB-RAT-TES-TAG-RAT-TAG
 - g'') PB-RAT-TAG-TES-RAT-TAG
 - k') PB-TES-TAG-RAT-TAG-RAT
 - k'') PB-TAG-TES-RAT-TAG-RAT
 - k''') PB-TAG-RAT-TES-TAG-RAT
 - k''''') PB-TAG-RAT-TAG-TES-RAT
 - l') PB-RAT-TES-TAG-RAT-TAG-PB'
 - l'') PB-RAT-TAG-TES-RAT-TAG-PB'
 - l''') PB-RAT-TA-RAT-TES-TAG-PB'
 - l''''') PB-RAT-TAG-RAT-TAG-TES-PB'

m') PB-(RAT-(TES)_m-TAG)_n-PB';

mit n = 3 bis 50, bevorzugt mit n = 3 bis 10 und m = 0 oder 1 unabhängig voneinander für jedes m,

n') PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''; mit m = 0 oder 1 für jedes m unabhängig voneinander, wobei mindestens ein m = 1 ist,

o') PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''-PB''''-RAT-(TES)_m-TAG-(TES)_m-PB''''';

mit m = 0 oder 1 für jedes m unabhängig voneinander, wobei mindestens ein m = 1 ist,

p') PB-RAT-(TES)_m-TAG-(TES)_m-PB'-(PB-RAT-(TES)_m-TAG-(TES)_m-PB')_n-PB-(TES)_m-TAG-(TES)_m-RAT-PB'

mit n = 0 bis 10 und mit m = 0 oder 1 für jedes m unabhängig voneinander, wobei mindestens ein m = 1 ist, (PB, PB', PB'', PB''', PB'''' = Primerbindungsstellen, wobei die Primerbindungsstellen bevorzugt unterschiedliche Sequenzen haben aber auch die gleiche Sequenz besitzen können.)

[0031] Die Herstellung der Nukleinsäurekonstrukte kann über zirkuläre Nukleinsäurekonstruktorstufen erfolgen, wobei bevorzugt mindestens ein Ende einer Nukleinsäure mit einem Marker und gegebenenfalls das andere Ende ebenfalls mit einem Marker bzw. Adapter modifiziert werden und dann die Zirkularisierung erfolgt. Bevorzugt besitzt mindestens einer der vorhandenen Marker oder Adapter auch eine Tagging-Enzymbindungsstelle (TES), besonders bevorzugt sind zwei TES vorhanden, insbesondere für das Tagging-Enzym EcoP15I, das auf beiden Seiten des zirkulären Konstruktes einen TAG abschneidet wodurch ein lineares Nukleinsäurekonstrukt im Sinne der vorliegenden Erfindung entsteht, dass zwei TAGs an beiden Enden enthält.

[0032] Weitere bevorzugte Nukleinsäurekonstrukte, die z. B. aus zirkulären Konstruktorstufen erzeugt werden können und die insbesondere für die gleichzeitige Quantifizierung beider Enden von Nukleinsäuremolekülen von Interesse sind, weil sie es erlauben, die Herkunft der TAGs vom gleichen Molekül zu erkennen, sind

q) TAG-TES-RAT-PB-RAT-TES

r) TAG-TES-RAT-PB-PB'-RAT-TES

s) PB''-TAG-TES-RAT-PB-RAT-TAG

t) PB''-TAG-TES-RAT-PB-RAT-TAG-PB' oder

u) PB''-TAG-TES-RAT-PB-PB'-RAT-TAG-PB'''

v) TAG-RAT-PB-RAT

w) TAG-RAT-PB-PB'-RAT

x) PB''-TAG-RAT-PB-RAT-TAG

y) PB''-TAG-RAT-PB-RAT-TAG-PB'

z) PB'-TAG-RAT-PB- PB'-RAT-TAG-PB'''

aa) PB-TAG-RAT-TAG

bb) PB-TAG-RAT-TAG-PB'

(PB, PB', PB'', PB''' = Primerbindungsstellen, wobei die Primerbindungsstellen bevorzugt unterschiedliche Sequenzen haben aber auch die gleiche Sequenz besitzen können.)

[0033] Alle in der vorliegenden Erfindungsbeschreibung dargestellten Nukleinsäurekonstrukte können neben den erfindungsgemäßen Domänenabfolgen weitere Sequenzabschnitte, wie z. B. Linker, Restriktionssites, Ligationsstellen etc. besitzen, die allerdings in der Regel nicht explizit in der Domänenabfolge der Konstrukte genannt werden.

[0034] Die Polynukleotiddomänen mit zufälliger Sequenz (RAT) sind zur Markierung von kennzeichnenden Polynukleotiddomänen (TAGs) geeignet. Unter einem „Marker“ werden folglich Polynukleotide verstanden, die eine Primerbindungsstelle und einen RAT enthalten, aber keinen TAG. Zudem können Marker weitere Oligonukleotiddomänen enthalten wie z. B. Tagging-Enzymbindungsstellen (TES) und Linker- und Ligationsstellen (LS). Eine Ligationsstelle ist das für eine Ligation vorgesehene Ende des Markers, das sich distal zur PB befindet.

[0035] Für die Herstellung der zur Amplifikation vorgesehenen Nukleinsäurekonstrukte ist es weiterhin sinnvoll, neben den Markern auch bestimmte Oligonukleotiddomänen in einem „Adapter“ zusammenzufassen. Adapter im Sinne der vorliegenden Erfindung sind immer mit einer Primerbindungsstelle versehen und können weitere Oligonukleotiddomänen insbesondere Tagging-Enzymbindungsstellen, Linker bzw. Ligationsstellen, aber keine RATs oder TAGs enthalten.

[0036] Die Marker und Adapter enthalten bevorzugt unterschiedliche Primerbindungsstellen, die bei den hochparallelen Sequenzierverfahren nach den Herstellerangaben bevorzugt verwendet werden sollen. Diese liegen an den distal zum TAG gelegenen Enden der Marker oder Adapter.

[0037] Die einzelnen Oligonukleotiddomänen, wie z. B. PB, TAG und RAT, bzw. die Marker oder Adapter können mit den nach dem Stand der Technik üblichen Verfahren zudem mit weiteren bekannten Modifikationen versehen sein, die z. B.

- das Binden an eine feste Phase ermöglichen, etwa durch Biotinylierung der Oligonukleotide und Bindung an eine mit Streptavidin versehene feste Phase, oder „Fänger-Sequenzen“ enthalten. Als „Fänger-Nukleotidsequenzen“ können Sequenzen genutzt werden, die z. B. an einer festen Phase gebunden sind und mit ihren korrespondierenden Gegensträngen, die sich in Lösung befinden, hybridisieren können,
- die Ligation an einer Seite eines Oligonukleotids verhindern, wie z. B. Amino-C7 Modifikationen oder „Reverse Nukleotide (Nukleotide, mit 3'-5'-Orientierung anstatt 5'-3'-Orientierung)“, wobei das vor Ligation zu schützende Ende das distal zum TAG gelegene Ende des Markers oder Adapters ist,
- die Ligation an einer Seite eines Oligonukleotids ermöglichen oder erleichtern, z. B. durch Phosphorylierung des 5' Endes des entsprechenden Oligonukleotids, oder z. B. durch die Erzeugung von überhängenden Restriktionsschnittstellen („sticky ends“),
- die Hybridisierung erleichtern bzw. die Dehybridisierung des Doppelstrangs erschweren, z. B. durch die Verwendung von „Locked nucleotide acids“.
- Restriktionsschnittstellen beinhalten

[0038] Diese Modifikationen können in die einzelsträngigen Oligonukleotide integriert werden. Adapter können mit komplementären Oligonukleotiden zu doppelsträngigen Oligonukleotiden hybridisiert werden. Marker können gegebenenfalls ebenso hergestellt werden, bevorzugt wird der Marker jedoch durch den Einsatz einer Polymerase, insbesondere durch das Klenow Fragment der DNA-Polymerase I wie zuvor beschrieben doppelsträngig hergestellt.

[0039] Mit Hilfe von Markern können Nukleinsäurefragmente mit einer Sequenz, die für eine in der Probe vorkommenden Nukleinsäure kennzeichnend ist (TAG), eindeutig gekennzeichnet werden, d. h. jede RAT-TAG Kombination ist mit großer Wahrscheinlichkeit einzigartig. Um eine solche eindeutige Kennzeichnung vorzunehmen, sollte die Nukleinsäuredomäne mit zufälliger Sequenz (RAT) bevorzugt lang genug sein, um im Verhältnis zur Anzahl der in der Probe vorkommenden Nukleinsäuren, einen Überschuss an Markern mit unterschiedlichen RATs bereitstellen zu können. Bevorzugt werden RATs enthaltende Marker mit mehr als der doppelten Menge bezogen auf die zu markierenden Nukleinsäurefragmente oder Nukleinsäuren eingesetzt. Im Einzelfall kann aber auch eine Markeranzahl ausreichen, die der zu bestimmenden Nukleinsäureanzahl entspricht, insbesondere wenn sich alle RATs garantiert in ihrer Sequenz unterscheiden. Bevorzugte Sequenzlängen für die in den Nukleinsäurefragmenten enthaltenen RATs sind zwischen 2 und 100 Nukleotiden, besser zwischen 4 und 50 Nukleotiden, besonders bevorzugte RATs besitzen eine Gesamtlänge zwischen 4 und 15 Nukleotiden. Die gesamte zufällige Sequenz kann dabei in einer Markerdomäne enthalten sein oder sich auf mehrere Markerdomänen bzw. auf Domänen unterschiedlicher Marker in einem Nukleinsäurekonstrukt aufteilen. Mit den beschriebenen Gesamtlängen der Nukleinsäuredomänen mit zufälliger Sequenz (RATs) lässt sich theoretisch eine große Anzahl von Markern mit unterschiedlichen RATs erzeugen, wobei das verwendete Markerensemble üblicherweise nur einen kleinen Teil der sich aus der Sequenzlänge ergebenden denkbaren RATs enthält, so dass die Wahrscheinlichkeit, dass zufällig sequenzidentische RATs in zwei Markern auftreten, sehr gering ist. Folglich kann die Sequenzlänge der RATs der Anzahl der zu markierenden, kennzeichnenden Nukleinsäurefragmente individuell angepasst werden. So werden z. B. für eine Genexpressionsanalysen, bei der die in einer Gewebeprobe enthaltenen mRNAs quantitativ bestimmt werden sollen, bevorzugt Markermischungen mit über 10^2 Markern, besonders bevorzugt zwischen 10^4 und 10^8 Markern eingesetzt.

[0040] Selbst wenn einzelne Marker tatsächlich identische RATs enthalten sollten, verfälschen diese die quantitative Bestimmung nur, wenn sie mit einem sequenzgleichen oder sequenzähnlichen TAG verknüpft werden sollten. Da alle individuellen TAG-RAT-Kombination für die individuellen, in der Probe vorkommenden und quantitativ zu bestimmenden Nukleinsäuren in einer zur Amplifikation vorbereiteten Mischung nur einmal vorhanden sein sollten, können sequenzidentische Kombinationen einer kennzeichnenden Polynukleotiddomäne (TAG) und einer Polynukleotiddomäne mit zufälliger Sequenz (RAT) folglich als Kopien einer kennzeichnenden Polynukleotiddomäne einer ursprünglich in der Probe vorhandenen Nukleinsäure identifiziert werden.

[0041] Da während der Amplifikation der Nukleinsäurekonstrukte auch Polymerisationsfehler auftreten können bzw. auch die Bestimmung der Nukleinsäuresequenz nicht frei von Fehlern ist, ist es vorteilhaft auch sequenzähnliche RAT-TAG-Kombinationen bei der Auswertung zu eliminieren. Was als „sequenzähnliche“ Kombinationen gilt, kann im Prinzip frei festgelegt werden.

[0042] Wie häufig solche Fehler auftreten hängt von der Länge der TAGs und RATs, den Amplifikationsbedingungen und der Sequenzierungsmethode ab. Als „sequenzähnliche“ Kopien werden bevorzugt RAT-TAG-

Sequenzen gesehen, die auf 10 Nukleotide maximal eine (1) Abweichung aufweisen, bevorzugt werden als sequenzähnliche Kopie RAT-TAG-Kombinationen definiert, die auf 15 oder 20 Nukleotide nicht mehr als eine Abweichung besitzen. So sollten z. B. sequenzähnliche RAT-TAG-Kombinationen mit einer Sequenzlänge von 40 Basenpaaren maximal 4 Sequenzabweichungen besitzen sofern unter fehlertolerierenden Bedingungen gearbeitet wird. Sofern Amplifikationsfehler und Sequenzierungsfehler weitgehend eliminiert werden können, ist es vorteilhaft die Fehlergrenze für „sequenzähnliche“ RAT-TAG-Kombinationen zu senken.

[0043] Abb. 1 zeigt ein Beispiel von vier RAT-TAG-Kombinationen mit einem gleichen TAG. Die Kombinationen 1, 2 und 3 sind unterschiedlich, die Kombination 4 ist eine Kopie von 3. Der TAG kommt in der Probe somit dreimal vor. Für eine Quantifizierung würde nur eine Kombination gezählt werden, z. B. 3 während Kombination 4 aus dem Datensatz eliminiert wird.

[0044] Die den TAG enthaltende Nukleinsäure liegt bevorzugt doppelsträngig, z. B. als cDNA oder DNA, vor und wird im Anschluss an das oder die doppelsträngigen Konstrukt(e) (Marker, Adapter) ligiert. Für eine lineare Amplifikation von RNA, kann die zu amplifizierende Nukleinsäure jedoch auch einzelsträngig vorliegen. Einzelsträngige RNA kann mithilfe von RNA-Ligase z. B. der T4 RNA-Ligase vor einer Amplifikation mit Markern versehen werden.

[0045] Im Folgenden werden einige Strategien zur Herstellung der Nukleinsäurekonstrukte aus doppelsträngiger DNA beschrieben. Aufgrund des modularen Charakters der einzelnen Verfahrensschritte zur Herstellung der Konstrukte können analog auch Nukleinsäurekonstrukte mit anderen Domänenabfolgen erzeugt bzw. gleiche Nukleinsäurekonstrukte mit einer alternativen Abfolge von Verfahrensschritten erhalten werden.

[0046] Die einzelnen Verfahrensschritte, die zur Herstellung der Nukleinsäurekonstrukte vorteilhaft sind, sind

- die Anbindung von einzel- oder doppelsträngigen Oligonukleotiden, insbesondere von Markern und Adaptern, oder der Nukleinsäuren an eine feste Phase, sowie deren Abspaltung
- die Ligation der einzelnen Marker, Adapter und der Nukleinsäuren und
- der Verdau von Nukleinsäuresträngen mit einem Restriktions- und/oder Tagging-Enzym

[0047] In einer bevorzugten Ausführungsform zur Erzeugung quantitativ zu bestimmender, doppelsträngiger Nukleinsäurekonstrukte enthaltend z. B. TAGs aus cDNA bzw. genomischer DNA, werden die doppelsträngigen Nukleinsäuren mit einem Marker oder Adapter versehen, der an eine feste Phase gebunden werden kann.

[0048] Das Binden an eine feste Phase hat den Vorteil, dass z. B. nicht ligierte oder nicht benötigte Nukleinsäuren nach einem Restriktionsverdau entfernt werden können oder benötigte Nukleinsäuren von der festen Phase abgenommen werden können. Die Bindung der Nukleinsäuren an eine feste Phase ist für manche Verfahren allerdings auch verzichtbar.

[0049] Bevorzugt sind Marker oder Adapter hierzu biotinyliert und können an eine Streptavidin-gekoppelte feste Phase gebunden werden. Die zu untersuchenden Nukleinsäuren eines Gemischs können mit solchen Markern oder Adaptern ligiert werden, die am entgegengesetzten Ende des zur Ligation vorgesehenen Endes (Ligationsstelle = LS) biotinyliert sind. Dies kann z. B. nach Verdau der zu quantifizierenden DNA, mit einem Restriktionsenzym an eine Restriktionsschnittstelle oder auch an glatte Enden der Nukleinsäure geschehen. Biotinylierte Marker oder Adapter können bei cDNA bevorzugt auch durch die Verwendung biotinylierter einzelsträngiger Oligonukleotide geschehen, die am 3'-Ende eine Poly-T-Sequenz besitzen und durch Erst- und Zweitstrangsynthese beginnend am 3' Ende der cDNA integriert werden (als 3'-Ende der cDNA wird das dem 3'-Ende der mRNA entsprechende Ende verstanden. Es ist durch eine Oligo-A/T-Domäne charakterisiert).

[0050] In einer bevorzugten Ausführungsform erfolgt der Verdau der Nukleinsäure (cDNA, DNA) mit einem oder mehreren häufig schneidenden Restriktionsenzymen. Häufig schneidende Restriktionsenzyme sind z. B. NlaIII, Hsp92II, FstI, BfaI, MaeI, XspI, HpyCH4IV, MaeII, Tail, TscI, AluI, TaqI, BfuCI, Bsp143I, BstENII, DpnII, Kzo9I, MboI, NdeI, Sau3AI, BstKTI oder Csp6I. Diese Enzyme erzeugen Nukleinsäurefragmente mit einer durchschnittlichen Länge von 200 bis 300 Basenpaaren. Die quantitativ zu bestimmenden Nukleinsäuren des Gemischs können vor oder nach diesem Verdau z. B. nach Ligation mit biotinylierten Markern oder Adaptern an eine feste Phase gebunden werden. Alternativ zum Verdau können die zu bestimmenden Nukleinsäuren des Gemischs vor Binden an die feste Phase durch Scheren zerkleinert werden. Nicht an die feste Phase gebundenen Fragmente können durch Waschen entfernt werden.

[0051] Die gebundenen Fragmente können bevorzugt durch Ligation erneut mit einem Marker oder Adapter, der nicht biotinyliert sein sollte, versehen werden. Wurden die quantitativ zu bestimmenden Nukleinsäuren zu-

vor mit einem Restriktionsenzym verdaut, so können Marker oder Adapter mit den der Restriktionsschnittstelle entsprechenden Enden versehen werden und mit den verdauten, quantitativ zu bestimmenden Nukleinsäuren ligiert werden. Das andere Ende der Adapter oder Marker kann vor Ligation geschützt werden z. B. durch den Einbau einer Amino C7-Modifikation im distal zur LS befindlichen Ende des Markers oder Adapters. Nach mechanischem Scheren können die quantitativ zu bestimmenden Nukleinsäuren nach Auffüllen oder Entfernen von überstehenden Einzelstrangenden („sticky ends“), z. B. mit dem Klenow Fragment, ebenfalls mit nicht-biotinylierten, passenden Markern oder Adaptern versehen werden.

[0052] Die quantitativ zu bestimmenden Nukleinsäuren des Gemischs liegen hiernach bevorzugt in der Form: PB-TAG-RAT-PB', PB-RAT-TAG-RAT oder PB-RAT-TAG-RAT-PB' vor und können z. B. mit PCR amplifiziert und z. B. mit hochparallelen Sequenziermethoden sequenziert werden.

[0053] Zur quantitativen Analyse der Genexpression reicht es in der Regel aus, die cDNA-Enden (Torres T. T. et al., Gene expression profiling by massively parallel sequencing, Genome Res. 2008 18: 172–177) als TAGs zu verwenden. Dazu werden mRNAs z. B. unter Verwendung eines biotinylierten Poly-T-Primers (PB1 = Adapter) in eine cDNA umgewandelt und an eine feste Phase gebunden. Danach erfolgt der Verdau mit einem häufig schneidenden Restriktionsenzym (Anchoring Enzyme). Im Anschluss erfolgt die Ligation mit einem PB'-RAT-Marker enthaltend eine zweite Primerbindungsstelle (PB2), worin das distal zur Ligationsstelle befindliche Ende vor Ligation geschützt ist. Das erhaltene Nukleinsäurekonstrukt ist folglich PB2-RAT-TAG-PB1. Die erhaltenen Nukleinsäurekonstrukte werden mittels PCR amplifiziert und im Anschluss sequenziert.

[0054] Eine weitere Möglichkeit für hochparallele Sequenzierverfahren besonders geeignete Nukleinsäurekonstrukte zu erhalten, ist die Verwendung von biotinylierten Markern oder Adaptern, deren Bindung an eine feste Streptavidin-Phase und die Gewinnung einzelsträngiger Konstrukte nach Denaturierung. Die Marker oder Adapter sind distal zur Primerbindungsstelle 5'-phosphoryliert und am Ende an dem die Primerbindungsstelle liegt vor Ligation geschützt, z. B. mit einer Amino-C7-Modifikation. Die für die Pyrosequenzierung besonders geeigneten Konstrukte c, d und e werden dadurch erhalten, dass der TAG, der z. B. eine cDNA bzw. ein genomisches DNA-Fragment sein kann, mit Markern und Adaptern ligiert wird, z. B. über spezifische Restriktionsschnittstellen oder glatte Enden (blunt-ends), wobei einer der Einzelstränge des aus zwei Einzelsträngen aufgebauten Markers, z. B. PB-RAT-LS oder Adapters, z. B. PB'-LS, biotinyliert ist. Nach einer Ligation entstehen zusätzlich die unerwünschten Nebenprodukte PB'-TAG-PB sowie PB-RAT-TAG-RAT-PB. Diese sind nun entweder an beiden Strängen biotinyliert oder überhaupt nicht biotinyliert. Nach Binden an eine Streptavidin-Phase werden zunächst die ungebundenen Fragmente durch Waschen entfernt. Denaturiert man nun die doppelsträngigen Konstrukte, so bleiben die Fragmente mit zwei biotinylierten Adaptern an der festen Phase. Lediglich diejenigen Nukleinsäurekonstrukte, mit der gewünschten Formation gehen als Einzelstrang in Lösung und können so gewonnen werden. Die einzelsträngigen Konstrukte können mit Hilfe einer Polymerase wieder doppelsträngig erzeugt werden. Dieses Verfahren ist vielseitig einsetzbar und kann im Prinzip für alle doppelsträngigen DNA-Nukleotide verwendet werden.

Gewinnung von TAGs durch Verwendung eines Tagging Enzyms:

1) Gewinnung eines mit einem RAT versehenen Tags einer cDNA:

[0055] In einer weiteren Ausführungsform werden die quantitativ zu bestimmenden Nukleinsäuren eines Gemischs wie zuvor beschrieben mit Nukleinsäurekonstrukten versehen (Markern und Adaptern), die eine Tagging-Enzym-Bindungsstelle (TES) enthalten und an einem Ende an ein Biotin-Molekül gebunden sind (z. B. TES-RAT-PB-Biotin). Damit lassen sich die Nukleinsäuren nun an eine feste Phase, z. B. Streptavidin-gekoppelte magnetische Partikel "magnetic beads" oder "beads", binden. Nach dem Verdau mit einem oder mehreren häufig schneidenden Restriktionsenzymen oder Scheren kann das Konstrukt nachdem es an eine feste Phase gebunden wurde mit einem Tagging-Enzym, wie z. B. BsmFI oder Mmel geschnitten werden.

[0056] An dieser Stelle sei erwähnt, dass Fragmente, die durch ein Tagging-Enzym abgespalten werden eine gleiche Länge haben, sofern die verwendeten Marker und Adapter ebenfalls von gleicher Länge waren und bei allen Molekülen die gleiche Abfolge von Domänen vorliegt, wie dies gemäß der vorliegenden Erfindung bevorzugt der Fall ist. Die durch das Tagging-Enzym abgespaltenen Nukleinsäurefragmente können somit einfach anhand der Größe, z. B. elektrophoretisch oder per HPLC, getrennt und isoliert werden. Auch nach einer Ligation mit einem Marker oder Adapter von bestimmter Größe können die Ligationsprodukte anhand

der Größe isoliert werden. Auch die entsprechenden spezifischen Amplifikationsprodukte können so isoliert werden.

[0057] Sofern ein TypIII-Enzym wie etwa EcoP15I als Tagging-Enzym verwendet wird, muss vorher noch ein Adapter oder Marker enthaltend eine zweite TES an das freie Ende der an die feste Phase fixierten Nukleinsäurestränge ligiert werden. Nach dem Verdau mit dem Tagging-Enzym und dem Entfernen der abgespaltenen, nicht an die feste Phase gebundenen Produkte können etwaige überstehende Enden z. B. mit dem Klenow Fragment, aufgefüllt bzw. mit einer Einzelstrang-spezifischen Exonuklease abgespalten werden, damit die Tagging-Enzym-Schnittstelle glatte Enden („blunt ends“) besitzt. Dann können Marker oder Adapter an den fixierten Nukleinsäurestrang gebunden werden. Alternativ dazu können mit „sticky ends“ versehene Adapter oder Marker verwendet werden, die dem durch das Tagging-Enzym entstehenden Ende komplementär sind.

[0058] Es entstehen bevorzugt, je nach Positionierung des Markers oder Adapters die Produkte
PB-TAG-TES-RAT-PB'
oder PB-RAT-TAG-TES-RAT-PB'

[0059] Nukleinsäurekonstrukte können sowohl unter Verwendung einer festen Phase als auch in Lösung hergestellt werden. In einer bevorzugten Ausführungsform werden dazu die in einer Probe enthaltenen Nukleinsäuren oder Nukleinsäurefragmente an beiden Enden mit Markern oder Adaptern ligiert. Die Enden können vorher durch ein Restriktionsenzym definiert werden und mit passenden Markern oder Adaptern versehen werden. Das erhaltene Konstrukt weist die folgende Domänenabfolge auf:

PB-Nukleinsäure-PB
oder:
PB-RAT-Nukleinsäure-RAT-PB

[0060] Verdaut man dieses Konstrukt mit einem häufig schneidenden Enzym, entstehen in vielen Fällen weitere Schnittstellen in der Nukleinsäure. Diese können für die Ligation eines weiteren Adapters oder Markers verwendet werden. Es entstehen die Produkte:

PB-RAT-TAG-PB
oder
PB-TAG-RAT-PB

[0061] Werden im ersten Schritt Adapter oder Marker eingesetzt, die Tagging-Enzym Erkennungsstellen enthalten, entsteht:

PB-RAT-TES-Nukleinsäure-TES-RAT-PB'
oder:
PB-TES-Nukleinsäure-TES-PB'

[0062] Danach können von den erhaltenen Nukleinsäuresträngen nach Verdau mit dem Tagging-Enzym jeweils zwei Nukleinsäurefragmente enthaltend einen TAG aus der Nukleinsäure abgespalten werden. Bevorzugt werden an die Schnittstellen der Tagging-Enzyme ein Marker oder Adapter angebunden. Es werden neben anderen folgende Nukleinsäurekonstrukte entstehen, die anhand ihrer Größe mittels z. B. Gelelektrophorese oder HPLC isoliert werden können:

PB-RAT-TES-TAG-PB'
und
PB'-RAT-TES-TAG-PB

oder:
PB-TES-TAG-RAT-PB'
und
PB'-TES-TAG-RAT-PB

[0063] Durch das Entstehen jeweils zweier Nukleinsäurekonstrukte je in der Probe enthaltener Nukleinsäure wird eine interne Kontrolle bei der quantitativen Bestimmung der Nukleinsäuren möglich.

[0064] Man erkennt, dass die zur Amplifikation benötigten Nukleinsäurekonstrukte, die die charakteristischen TAG-RAT-Kombinationen enthalten, auf vielfältige Weise erzeugt werden können. Der Aufbau der Nukleinsäurekonstrukte kann dabei in modularer Weise erfolgen, so können z. B. die einzelnen Verfahrensschritte zur Anbindung bzw. Erzeugung einzelner Polynukleotide mit den unterschiedlichen funktionellen Domänen (TAG, RAT, PB, TES) einfach kombiniert werden, um die gewünschten Nukleinsäurekonstrukte zu erhalten. Dadurch

ist das erfindungsgemäße Verfahren vielfältig einsetzbar und nicht auf eine bestimmte Prozessierung der in der Probe enthaltenen Nukleinsäuren beschränkt.

[0065] Die einzelnen, modular anwendbaren Verfahrensschritte werden im Folgenden beispielhaft näher erläutert.

1. Vorbereitung einer RNA-haltigen Probe zur Herstellung von cDNA-TAGs:

[0066] Mit Hilfe eines handelsüblichen Kits z. B. mit Trizol reagent (Invitrogen Corp., Carlsbad, USA) wird Gesamt-RNA aus einer biologischen Probe gewonnen. Aus der Gesamt-RNA wird mit einem handelsüblichen Kit, z. B. "Oligtex-Midi-Kit" (Qiagen N. V., Venlo, Niederlande) die mRNA isoliert. Die cDNA wird durch reverse Transkription unter Verwendung eines 5'-biotinylierten Poly-T-Oligonukleotides, z. B. mit dem „cDNA synthesis system“, (Invitrogen Corp.), hergestellt. Das Produkt wird zu einer doppelsträngigen DNA konvertiert („cDNA synthesis system“, Invitrogen Corp.) und in einem geeigneten Puffer gelöst. In einer besonders bevorzugten Variante des Verfahrens soll der verwendete Poly-T-Oligonukleotid die Erkennungssequenz für EcoP15I CA-GCAG bzw. CTGCTG enthalten z. B. 5'-Biotin-PB-CAGCAGTTTTTTTTTTTTTTTTTTT-3' (SEQ ID No. 1).

2. Enzymatischer Verdau von in der Probe vorhanden Nukleinsäuren:

[0067] Doppelsträngige DNA wird in einer Reaktionslösung mit einem häufig schneidenden Restriktionsenzym („Anchoring-Enzym“), z. B. NlaIII, in gepufferter Lösung verdaut. Nach dem Verdau wird die erhaltene DNA extrahiert, mit Alkohol präzipitiert und in einem geeigneten Puffer aufgenommen. Hiernach wird die DNA-Lösung mit Streptavidin-beschichteten Partikeln versetzt. Die Partikel werden bei Raumtemperatur in geeignetem Puffer inkubiert, um eine Bindung der biotinylierten cDNA mit den Streptavidin-beschichteten Partikeln zu ermöglichen. Die an die Partikel gebundene DNA wird gewaschen und in einem geeigneten Puffer aufgenommen.

3) Herstellung und Ligation von Polynukleotid-Domänen, Adaptern und Markern:

[0068] Adapter werden aus zwei komplementären einzelsträngigen Oligonukleotiden zusammengesetzt. Die beiden einzelsträngigen Oligonukleotide werden in einem geeigneten Puffer gelöst und miteinander zu einem Doppelstrang hybridisiert. Der Marker enthaltend den RAT wird ebenfalls zunächst einzelsträngig synthetisiert und von dieser Matrize wie zuvor beschrieben mit Hilfe des Klenow Fragments ein Doppelstrang synthetisiert. Adapter und Marker sind vor einer ungewollten Ligation auf der distal zur LS befindlichen Seite durch den Einbau von z. B. einer Amino-C7-Modifikation der 3' Seite des entsprechenden Oligos geschützt, während an der Ligationsstelle durch 5'-Phosphorylierung die Ligation ermöglicht wird. Der doppelsträngige Marker oder Adapter kann nun mit einer Ligase, z. B. T4-Ligase (Invitrogen Corp.) in einem geeigneten Puffer kovalent mit einem 3'-Ende eines anderen Polynukleotids bzw. einer Nukleinsäure aus der Probe oder eines Fragmentes davon, das den TAG enthält und an eine feste Phase gebunden sein kann, verknüpft werden. Sofern Marker und Adapter ligiert werden sollen, werden diese bevorzugt in einer ausreichend großen Menge zugegeben, so dass alle freien Enden der aufzubauenden Nukleinsäurestränge auch mit den Polynukleotiden ligiert werden.

4) Schneiden mit dem Tagging Enzym:

[0069] In einem geeigneten Puffer werden die mit einem eine TES enthaltenden Adapter oder Marker versehenen, zu quantifizierenden Nukleotide mit einem Tagging-Enzym, z. B. BsmFI, MmeI, EcoP15I, inkubiert. Die abgespaltenen Nukleinsäurefragmente werden entweder nach Elektrophorese oder HPLC isoliert, oder sind (bei Einsatz entsprechender Marker oder Adapter) an eine feste Phase gebunden. Sollen an die Tagging-Enzym-Schnittstelle weitere Marker bzw. Adapter angebunden werden und ist die Schnittstelle keine glatte Schnittstelle, kann das überhängende Ende aufgefüllt bzw. mit einer Exonuklease abgespalten werden. Dazu kann z. B. das Klenow-Fragment der DNA-Polymerase I verwendet werden.

5) Amplifikation:

[0070] Zur Vervielfältigung der Nukleinsäurekonstrukte werden bevorzugt PCR-basierte Amplifizierungsverfahren verwendet. Alternativ können auch eindirektionale Amplifizierungs- oder Klonierungsverfahren zur Vervielfältigung der Nukleinsäurekonstrukte eingesetzt werden.

6) Sequenzierung:

[0071] Die Sequenz individueller, durch Amplifikation erhaltener Nukleinsäurefragmente in einer Mischung wird bevorzugt durch geeignete Sequenzierverfahren bestimmt. Solche Verfahren werden kommerziell angeboten, z. B. „454“-Picoliter Verfahren, mit dem GSFLX-System (Roche Diagnostics, Deutschland,), Solexa-Verfahren, (Illumina Inc., San Diego USA); SOLid-Verfahren (Applied Biosystems Inc, Foster City; USA); Heliscope™ Single Molecule Sequencer, (Helicos Corp., Cambridge, USA). Mit diesen Methoden können heute schon bis zu mehreren Millionen Sequenzen parallel sequenziert werden. Dabei kann eine Mischung aus amplifizierten Nukleinsäurekonstrukten ohne weitere Auftrennung bzw. ohne zwischengeschalteten Klonierungsschritt parallel sequenziert werden. Allerdings ist auch möglich, die erhaltenen Nukleinsäurefragmente zu Konkatemenen zusammenzufügen, die im Anschluss kloniert, aufgearbeitet und mit herkömmlichen Verfahren (z. B. nach Sanger) sequenziert werden können.

[0072] Da die quantitative Auswertung der erzeugten Nukleinsäurekonstrukte durch Sequenzierung erfolgt, sollte der zu sequenzierende Teil der Konstrukte in einer bevorzugten Ausführungsform eine Länge von 400 Nukleotiden nicht übersteigen. Bevorzugte Nukleinsäurekonstrukte besitzen daher eine Länge zwischen 15 und 150 Basenpaaren, besonders bevorzugt zwischen 25 und 75 Basenpaaren. Allerdings können auch längere Sequenzen bis hin zu den gesamten Nukleinsäuresequenzen sequenziert werden.

[0073] Kurze Beschreibung der Abbildungen, die beispielhaft einzelne mögliche Ausführungsformen der Erfindung verdeutlichen:

[0074] **Abb. 1** zeigt ein Beispiel von vier RAT-TAG-Kombinationen mit einem gleichen TAG.

[0075] **Abb. 2** zeigt schematisch die Herstellung von Nukleinsäurekonstrukten aus mRNA zur quantitativen Untersuchung der Genexpression.

[0076] In **Abb. 2**, Schritt 1.1 bis 1.3 wird die lineare Amplifikation unter Verwendung eines T7-Promotors (PB') gezeigt. Dazu wird ein Marker mit der Domänenabfolge PB'-PB-RAT-Poly-T, wobei PB' die Promotor-Sequenz für eine T7- oder SP6-RNA-Polymerase enthält, mit dem Poly-A-Ende der mRNAs in der Probe hybridisiert. Das freie 3'-Ende der Poly-T-Sequenz dient dann als Startpunkt für eine reverse Transkription (Erststrangsynthese: **Abb. 2**, Schritt 1.1). Durch Zweitstrangsynthese z. B. unter Verwendung von Oligohexameren wird daraus eine doppelsträngige cDNA hergestellt (Schritt 1.2). Aus der cDNA kann dann unter Verwendung der T7-Polymerase eine lineare Amplifikation zur Herstellung von antisense RNA (aRNA) erfolgen (Schritt 1.3). Die erhaltene aRNA kann dann unter Verwendung von „random Hexamers“ als Primer durch reverse Transkription in den entsprechenden DNA-Erststrang umgeschrieben werden (Schritt 1.4), es wird die entsprechende amplifizierte, einzelsträngige cDNA erhalten. Mit Hilfe der Primerbindungsstelle (PB) kann dann der entsprechende DNA-Erststrang amplifiziert werden (Schritt 1.6.1), wobei auch biotinylierte Primerbindungsstellen verwendet werden können (Schritt 1.6.2). Die erhaltenen doppelsträngigen Konstrukte können dann z. B. wie in **Abb. 3** (Schritt 3–5) gezeigt weiter analysiert werden. In **Abb. 2**, Schritte 2–4, wird eine Amplifikation von mRNA mit dem sogenannten SMART-Verfahren gezeigt.

[0077] **Abb. 3** zeigt schematisch die Herstellung von Nukleinsäurekonstrukten aus genomischer DNA am Beispiel des Digital Karyotyping. Dazu wird in einer Probe die vorliegende genomische DNA unter Verwendung eines methylierungsinsensitiven Restriktionsenzym verdaut (**Abb. 3**, Schritt 1), die gewonnenen Fragmente besitzen bevorzugt überhängende Enden, so dass an die entsprechenden Restriktionsschnittstellen der genomischen DNA Marker der Form Biotin-PB-RAT ligiert werden können, welche passende Ligationssstellen distal zur Primerbindungsstelle aufweisen (Schritt 2). Die erhaltenen Nukleinsäure-konstruktstufen können dann mit einem weiteren Restriktionsenzym verdaut werden (Schritt 3). Nach Anbindung der Marker-haltigen Fragmente an eine Streptavidin-Matrix (z. B. Streptavidin-beschichtete magnetische Partikel), kann dann ein Adapter enthaltend eine zweite Primerbindungsstelle (PB') an die gebundenen Fragmente über die entsprechenden überhängenden Enden („sticky ends“) ligiert werden. Die so erhaltenen Nukleinsäurekonstrukte können dann amplifiziert und sequenziert werden. Beim methylierungsspezifischen DK wird das erste Enzym durch ein methylierungssensitives Enzym ersetzt. In einer alternativen Ausführungsform, die in **Abb. 4**, Schritt 5–8 gezeigt wird, können auch Tagging-Enzym-Bindungsstellen enthaltende Marker, die im vorliegenden Fall biotinyliert sind, und Adapter für den Aufbau der Nukleinsäurekonstrukte verwendet werden. Es werden dann die in Schritt 5 gezeigten Nukleinsäurekonstrukte erhalten. Durch Zugabe des Tagging-Enzyms werden dann kurze TAGs aus der Nukleinsäure geschnitten (Schritt 6). Sofern diese überstehende Enden aufweisen werden die erhaltenen Fragmente, dann z. B. mit dem Klenow-Fragment aufgefüllt. Nun können die in Lösung befindlichen adapterhaltigen Fragmente mit Markern zu dem entsprechenden Nukleinsäurekonstrukt über dessen glatte

Enden („blunt-ends“) ligiert (Schritt 7.1) und im Anschluss amplifiziert werden. Das amplifizierte Nukleinsäurekonstrukt kann dann z. B. elektrophoretisch gereinigt und sequenziert werden. Auch das an der festen Phase gebundene Fragment kann zu dem gewünschten Nukleinsäurekonstrukt weiterverarbeitet werden, indem man das glatte Ende mit einem Adapter enthaltend eine zweite Primerbindungsstelle ligiert (Schritt 7.2). Das gebundene Nukleinsäurekonstrukt kann direkt amplifiziert und im Anschluss sequenziert werden. Es können aber auch die in Lösung befindlichen Fragmente enthaltend den Adapter und einen TAG und die gebundenen Fragmente enthaltend den Marker und einen TAG miteinander zu Dtags ligiert werden (Schritt 6.2) bevor die Amplifizierung und Sequenzierung des erhaltenen Nukleinsäurekonstruktes erfolgt.

[0078] Die im Folgenden beschriebenen Anwendungen der Erfindung belegen die vielfältigen Verwendungsmöglichkeiten der Erfindung. Im Detail beschrieben werden:

- 1) RAT-Markierung von Nukleinsäuren vor ihrer Vermehrung durch Lineare Amplifikation:
- 2) Konstrukte für die Quantifizierung von cDNA,
 - 2.1.) Konstrukte für die Quantifizierung von 3'-Enden von cDNA,
 - 2.2.) Konstrukte für die Quantifizierung von 5'-Enden von cDNA,
 - 2.3.) Konstrukte für die gleichzeitige Quantifizierung von 5'- und 3'-Enden von cDNAs,
 - 2.4.) Konstrukte für die gleichzeitige Quantifizierung von 3'- und 5'-Enden der gleichen cDNA mit Hilfe einer Zirkularisierungsmethode,
- 3) Konstrukte für die Quantifizierung von cDNA, die nicht aus Poly-A-haltiger RNA hergestellt wurde,
- 4) Konstrukte für die Quantifizierung von Fragmenten genomischer DNA,
 - 4.1.) Konstrukte für die Quantifizierung von zuvor mit Bisulfit umgesetzter genomischer DNA,
 - 4.2.) Konstrukte für die Quantifizierung beider Enden von Fragmenten genomischer DNA mit Hilfe einer Zirkularisierungsmethode.

- 1) RAT-Markierung von Nukleinsäuren vor ihrer Vermehrung durch lineare Amplifikation:

[0079] Ziel der „linearen Amplifikation“ ist es, Nukleinsäuren so zu vermehren, z. B. um genügend Material für eine Genexpressionsstudie zu erhalten, dass die verhältnismäßige Zusammensetzung einer Nukleinsäure-Probe bewahrt bleibt. Hierzu existieren verschiedene Methoden. So wird etwa zunächst, z. B. aus einer mRNA, an die ein einzelsträngiger Marker enthaltend einen RAT, gebunden wurde, antisense RNA (aRNA) mit Hilfe der T7- oder SP6-RNA-Polymerase (US 6,916,633) oder mit Hilfe eines RNA-Primers, der einen T-7-Promotor Dafforn et al., Biotechniques. 2004 Nov; 37(5): 854–7 linear hergestellt. Jedoch treten auch bei der „linearen Amplifikation“ von RNA sequenzbedingte Abweichungen der Kopienzahl auf (e. g. Caretti et al 2008, J. Cellular Biochemistry 103: 556–563), die durch Einbindung eines einzelsträngigen RAT erkannt werden können. In einem alternativen Verfahren kann z. B. ein einzelsträngiger Marker enthaltend einen RAT an das 5' Ende von in der Probe enthaltenen mRNAs angebunden werden. Im Anschluss erfolgt dann die lineare Amplifikation z. B. vom 3'-Ende aus mit Poly-T Primern unter Verwendung einer reversen Transkriptase.

2) Konstrukte für die Quantifizierung von cDNA

[0080] Die lineare Amplifikation ist keine notwendige Voraussetzung für die Quantifizierung von cDNAs, die aus mRNA mit einem 3'-Poly-A-Ende hergestellt wurden. Sie bietet sich jedoch an, wenn die mRNA nur in geringen Mengen vorhanden ist. Daher werden zunächst Konstrukte beschrieben, welche die lineare Amplifikation der cDNA über einen RNA-Schritt ermöglichen. Im Weiteren werden dann Verfahren beschrieben, die von in genügend großen Mengen vorliegenden Nukleinsäuren ausgehen, seien diese durch Amplifikation erzeugt oder nicht. Danach werden beispielhaft Verfahren beschrieben, welche die Quantifizierung von kennzeichnenden Nukleinsäuredomänen (TAGs) aus Nukleinsäuren unter Verwendung von RATs ermöglichen.

2.1. Konstrukte für die Quantifizierung von 3'-Enden von cDNA

[0081] Zur Amplifikation wird die mRNA zunächst mit Hilfe eines Oligo-dT-haltigen Primers in cDNA umgewandelt. Für die spätere Amplifikation mit einer Polymerase, wie z. B. der SP6-, T3- oder T7-RNA-Polymerase, enthält dieser Primer auch die Promotorsequenz für die Polymerase. Die spätere, doppelsträngige cDNA hat dann die Domänenstruktur „cDNA-Oligo-A/T-Promotor“. Veränderungen der Nukleinsäure-Zusammensetzung können erkannt werden, wenn zwischen dem Promotor und der zu amplifizierenden DNA ein RAT integriert wird. Die resultierende Domänenstruktur ist in diesem Fall

„cDNA-Oligo-A/T-RAT-Promotor“,

wobei die T7-Promotor-Sequenz auch gleichzeitig als Primerbindungsstelle dienen kann. Es kann allerdings auch eine separate Primerbindungsstelle (PB) neben dem Promoter (Polymerasebindungsstelle) eingebaut werden, so dass die folgende Domänenstruktur erhalten wird:

"cDNA-Oligo-A/T-RAT-PB-Promotor",

[0082] Für die Quantifizierung ohne vorhergehende direkte Amplifikation z. B. mit Hilfe einer Polymerase, wie z. B. der SP6-, T3 oder T7-Polymerase wird ein Konstrukt gewählt, das statt des jeweiligen Promotors eine Primer-Bindungsstelle (PB) und/oder eine TES enthält. Eine entsprechende Domänenstruktur ist dann: cDNA-Oligo-A/T-TES-RAT-PB

[0083] Die beiden folgenden bevorzugten Verfahren können ebenfalls verwendet werden, um TAGs zu quantifizieren, die von den 3'-Enden einer cDNA gewonnen wurden, die den Poly-A-Teil enthalten.

[0084] Für die Quantifizierung von cDNA-Enden in einem bevorzugten Verfahren wird mRNA mit einem Poly-T enthaltenden Marker Poly-T-RAT-PB-Biotin mittels Erst- und Zweitstrangsynthese in cDNA umgeschrieben. Die cDNA wird danach mit einem oder mehreren häufig schneidenden Enzymen (Anchoring Enzyme) verdaut. Die Marker-haltigen Fragmente werden dann an eine Streptavidin-modifizierte Matrix wie z. B. magnetische Partikel gebunden. Die nicht gebundenen Fragmente werden durch Waschen entfernt. Es folgt die Ligation mit einem Adapter enthaltend eine Ligationsstelle am distal zur Primerbindungsstelle gelegenen Ende, die durch eine Amino-C7-Modifikation vor Ligation geschützt ist, wobei die überhängenden Enden für die Ligation genutzt werden. Nach Entfernen der nicht ligierten Adapter können die erhaltenen Nukleinsäurekonstrukte amplifiziert und sequenziert werden. Alternativ kann auch in dem oben beschriebenen Verfahren anstelle des Adapters ein Marker verwendet werden. In diesem Fall ist es nicht notwendig, die mRNA mit einem RAT enthaltenden Poly-T-Oligonukleotid umzuschreiben.

[0085] Eines der bevorzugten Verfahren zur quantitativen Untersuchung der Genexpression wird als „Super-TAG“ Verfahren bezeichnet. Dazu wird mRNA unter Verwendung eines 3' biotinylierten Poly-T Oligonukleotids, mit einer EcoP15I-Bindungsstelle (EcoP15ITES) vor der Poly-T Sequenz in cDNA umgeschrieben. Danach erfolgt der Verdau mit einem häufig schneidenden Restriktionsenzym (Anchoring Enzyme), z. B. mit NlaIII. Die geschnittene cDNA wird durch die Biotin-Gruppe an Streptavidin-beschichtete magnetische Partikel gebunden. Alternativ können die Oligo-dT bereits vor der cDNA-Synthese an die Partikel gebunden sein oder der Verdau mit dem Anchoring Enzyme erst nach Binden an die Partikel stattfinden. Die nicht gebundenen Fragmente werden durch Waschen entfernt. Im nächsten Schritt erfolgt die Ligation mit einem doppelsträngigen PB'-EcoP15 ITES Oligonukleotid (Adapter). Die erhaltenen Nukleinsäurestränge werden dann mit EcoP15I inkubiert und die abgespaltenen Fragmente nach einer Gelelektrophorese isoliert. Die Adapter-TAG Konstrukte können nun mit einem Marker ligiert werden. Die erhaltenen Nukleinsäurekonstrukte besitzen die folgende Domänenabfolge: PB'-EcoP15ITES-TAG-RAT-PB.

[0086] Im Anschluss erfolgen eine PCR-Amplifikation und die Sequenzierung der amplifizierten Konstrukte.

2.2. Konstrukte für die Quantifizierung von 5'-Enden von cDNA

[0087] Für die genaue Quantifizierung der 5'-Enden von cDNAs können RATs an der 5' CAP-Sequenz der mRNA angebracht werden. Die 5'-CAP-Sequenz besteht aus einem Guanin-Nukleotid, das durch eine ungewöhnliche 5'- zu 5'-Triphosphat-Bindung an die mRNA gebunden ist. Zur Anbindung eines RAT kann z. B. analog zur Herstellung von Voll-Längen-cDNA verfahren werden (Clepet et al. Improved full-length cDNA production based on RNA tagging by T4 DNA-Ligase. Nucleic Acids Research, 2004, Vol. 32, No. 1 e6). Dabei werden RNAs z. B. mit „Alkaline Phosphatase“ behandelt. Transkripte mit intakter 5'-CAP-Sequenz werden von diesen Enzymen nicht dephosphoryliert, während gebrochene mRNA-Moleküle und nicht-codierende RNA, die eine solche CAP-Sequenz nicht besitzen, dephosphoryliert werden und statt der Phosphat-Gruppe mit einer 5'-OH Gruppe enden, an welche die RNA-Polymerase keine RNA-Oligonukleotide anbinden kann. Im zweiten Schritt wird dann die CAP-Sequenz enzymatisch entfernt, wobei die 5'-Phosphatgruppe frei wird, die dann für die Ligation eines RNA-Oligonukleotids zur Verfügung steht. Dieses Oligonukleotid bekannter Sequenz dient später als Primerbindungsstelle (PB) für die Zweitstrang-Synthese der cDNA. Um in dieses Konstrukt einen RAT einzufügen, sollte die Sequenz des RNA-Oligonukleotids so gewählt werden, dass das spätere RNA-Konstrukt die Domänenfolge 5'-PB-RAT-mRNA-3' besitzt. Wahlweise kann in dieses Konstrukt eine (T7-Polymerase-) Promotorsequenz eingefügt werden, die dann als Polymerasebindungsstelle und gleichzeitig als Primer-Bindestelle (PB) dienen kann. Die entsprechende Domänenstruktur ist dann '5'-(T7)Promotor-RAT-mRNA-3'

[0088] Eine andere Methode um das 5' Ende einer RNA mit einem Primer zu versehen, ist die Verwendung des „Switch Mechanism At the 5'-end of Reverse Transcript“ (SMART, TAKARA, Seta Otsu, Japan). Hierbei entsteht bei der Erststrang-Synthese der cDNA ein von der Reversen Transkriptase gebildeter, aus mehreren

(meist drei) Cytosin-Basen bestehender Überhang auf der 3'-Seite des neu entstandenen DNA-Strangs (siehe dazu auch **Abb. 2**, Schritte 2 bis 4). Dieser steht als Bindestelle für einen Oligonukleotid-Primer zur Verfügung. Hier kann z. B., um den RAT einzufügen, ein Oligonukleotid, das am 5'-Ende eine Primerbindestelle besitzt, gefolgt von einem RAT der wiederum mit drei aufeinanderfolgenden Guanin-Basen am 3'-Ende abschließt als Primer für die Zweitstrangsynthese der cDNA verwendet werden.

[0089] Für viele Zwecke reicht es aus, lediglich einen TAG vom 5'-Ende einer cDNA zu gewinnen, der dann zur Quantifizierung der ursprünglichen mRNA dienen kann. Dazu wird wie oben beschrieben ein Nukleinsäurekonstrukt enthaltend zumindest eine PB (oder eine Promotorsequenz, wie z. B. von SP6, T3 oder T7), einen RAT und eine TES an die nach der Dephosphorylierung und Entfernung der CAP-site frei werdende 5'-Phosphatgruppe gebunden. Das entsprechende einzelsträngige RNA-Konstrukt hat dann bevorzugt die Form 5'-Biotin-PB/Promotor-RAT-TES-mRNA.

[0090] Gleichzeitig kann an die Poly-A-Sequenz des 3'-Ende der mRNA ein DNA-Konstrukt, das eine Poly-T-Sequenz enthält, durch „Wasserstoffbrückenbildung“ zwischen der Poly-A- und der Oligo-T-Sequenz des Konstrukts gebunden werden. Dieses Konstrukt enthält zumindest eine Oligo-T-Domäne, kann aber auch aus weiteren Domänen bestehen (siehe unten). Ausgehend von diesem doppelsträngigen Ende erfolgt dann die Reverse Transkription, an die sich die Zweitstrangsynthese anschließt. Das entstehende, doppelsträngige Produkt hat dann die Domänenstruktur

5'-Biotin-PB/Promotor-RAT-TES-cDNA-Oligo-A/T-3'

und kann über die Biotin-Gruppe an eine feste Phase gebunden werden. Wenn die TES die Erkennungssequenz eines Typ-II-Restriktionsenzym handelt, kann das Produkt sofort damit geschnitten werden. Während der nicht gebundene Teil der cDNA weggewaschen werden kann, bleibt an die feste Phase ein Konstrukt mit der Domänen-Abfolge

5'-Biotin-PB/Promotor-RAT-TES-TAG.-3'

gebunden, wobei der TAG, der in diesem Fall eine definierte Länge hat, die cDNA charakterisiert.

[0091] Alternativ dazu kann ein TAG auch im Sinne der Erfindung durch mechanische Zerkleinerung der cDNA oder Behandlung mit einer unspezifischen Endo- oder Exonuklease gewonnen werden. Er besitzt dann allerdings nicht per Definition eine definierte Länge. In diesem Fall entfällt die Notwendigkeit, eine TES-Domäne in das Konstrukt einzuführen. Es reicht die Domänen-Abfolge

5'-Biotin-PB/Promotor-RAT-TAG,

um die cDNA zu charakterisieren.

[0092] An die nach Verdau oder Zerkleinerung frei werdenden 3'-Enden kann eine weitere Primerbindestelle (PB), ein Adapter oder Marker gebunden werden. Dieses Konstrukt, das im Falle der Verwendung eines Tagging-Enzyms eine definierte Länge hat, kann dann durch Amplifikation, ausgehend von den beiden darin enthaltenen PBs, vermehrt werden.

[0093] Wenn es sich bei dem Tagging-Enzym um eine TypIII-Endonuklease wie etwa EcoP15I handelt, wird an die Schnittstelle des Restriktionsenzym bzw. an das durch die unspezifische Endonuklease oder die mechanische Zerkleinerung entstehende freie Ende, das distal zur Biotin-Gruppe liegt, bevorzugt ein Konstrukt, das eine PB plus eine TES enthält so angefügt, dass ein Konstrukt der Form

5'-Biotin-PB/Promotor-TAG-TES-PB-3'

entsteht. Zusätzlich kann auch hier ein RAT eingefügt werden. Das Konstrukt hat dann die Form:

5'-Biotin-PB/Promotor-TAG-TES-RAT-PB-3'

[0094] Solche Konstrukte können mittels PCR amplifiziert, anhand ihrer definierten Größe identifiziert und isoliert werden (Gelektrophorese, HPLC).

[0095] In einer besonders bevorzugten Ausführungsform des Herstellungsverfahrens werden als feste Phase bevorzugt Streptavidin-beschichtete, magnetische Partikel verwendet. Auch Streptavidin-beschichtete Reaktionsgefäße sind eine bevorzugte Alternative. Das Partikel-gebundene Ende der erzeugten Nukleinsäurestränge ist gegen eine Ligationsreaktion zu schützen.

2.3.) Konstrukte für die gleichzeitige Quantifizierung von 5'- und 3'-Enden von cDNAs

[0096] Oben beschrieben Konstrukte von der Form:

5'-Biotin-PB/T7-Promotor-RAT-TES-cDNA-Oligo-A/T-3' oder andere, schon beschriebene Konstrukte erlauben es prinzipiell, lediglich einen TAG von einem der cDNA-Enden zu gewinnen. In speziellen Fällen wird jedoch

die Gewinnung zweier TAGs vom 5'- und 3'-Ende der gleichen cDNA gewünscht, die damit als solche erkannt werden können. Dazu wird bevorzugt ein zirkuläres Produkt der Form

5'-PB-cDNA-Oligo-A/T-PB'-3'

hergestellt, wobei PB und PB' bevorzugt nicht die gleiche Sequenz haben. Nachdem PB und PB' an die Enden der cDNA gebunden wurden, können die so mit bekannten Sequenzen versehenen cDNA-Enden durch Ligation ihrer PB-Enden unter geeigneten Bedingungen zu einem zirkulären Molekül umgestaltet werden. Diese zirkulären Moleküle können jetzt durch PCR ausgehend von den PB-Domänen vermehrt werden. Um die ursprüngliche Zusammensetzung der cDNA-Population mit Hilfe von RATs ermitteln zu können, kann mindestens ein Ende mit einem Marker versehen werden der eine PB-Domäne enthält. Dies kann über die CAP-Site und/oder das Poly-A-Ende erfolgen, wie oben beschrieben, wobei vor der Zirkularisierung an das 5'-Ende der cDNA nach dem oben beschriebenen Entfernung der CAP-Site ein Konstrukt der Form

5'-PB-RAT-cDNA-3'

gebildet wird. Am 3'-Ende wird ein Konstrukt der Form

5'-cDNA-Oligo-A/T-PB'-3'

oder

5'-cDNA-Oligo-A/T-RAT'-PB'-3'

gebildet. PB und PB' können dabei sowohl komplementäre, überhängende Einzelstrangenden haben, die sich besonders für eine Zirkularisierung eignen, als auch glatte Enden besitzen. Die Ligation zur Zirkularisierung der Moleküle erfolgt bevorzugt in genügend großer Verdünnung, so dass die Wahrscheinlichkeit, dass ein Ende des gleichen cDNA-Moleküls dem anderen Ende näher ist als ein beliebiges Ende eines anderen Moleküls, ausreichend groß ist. Aus den so zirkularisierten Molekülen können dann entweder enzymatisch oder durch mechanische Zerkleinerung Konstrukte gewonnen werden, die einen TAG von jedem Ende des Moleküls enthalten. Sie haben dann die Form:

5'-TAG-RAT-PB-PB'-RAT'-TAG-3'

[0097] Durch Anfügen zweier PBs an die Enden dieses Konstrukts entstehen Konstrukte der Form

1) 5'-PB"-TAG-RAT-PB-PB'-RAT'-TAG-PB"'-3',

oder, je nachdem welches 3'-Konstrukt verwendet wurde

2) 5'-PB"-TAG-RAT-PB-PB'-TAG-PB"'-3',

das ausgehend von PB" und PB"' amplifiziert und dann sequenziert werden kann, wobei PB" und PB"' bevorzugt aber nicht notwendigerweise ungleiche Sequenzen haben.

[0098] Um Konstrukte definierter Länge zu erhalten, kann eine Größenselektion z. B. über HPLC oder Gelelektrophorese erfolgen, oder es können ein oder mehrere Tagging-Enzyme verwendet werden.

[0099] Für die Verwendung von Typ-II-Enzymen wird das 5'-Ende der cDNA mit einem Konstrukt der Form

5'-TES-(CAP-Site)-cDNA-3'

oder

5'-PB-TES-(CAP-Site)-cDNA-3'

versehen, wobei die TES bevorzugt zusammen mit der PB Teil eines Adapters ist.

3) Konstrukte für die Quantifizierung von cDNA, die nicht aus Poly-A-haltiger RNA hergestellt wurden

[0100] Viele RNA-Moleküle in eukaryotischen Zellen sowie nahezu alle bakteriellen und viralen RNAs enthalten, anders als Protein-codierende eukaryotische mRNAs, keine 3'-Poly-A-Sequenz. Um sie zu quantifizieren, werden sie bevorzugt zunächst mit einem einzel- oder doppelsträngigen RNA-Oligonukleotid bekannter Sequenz ligiert, das später als Primerbindestelle dienen kann. Einzelsträngige Oligonukleotide können mit Hilfe von Einzelstrang-RNA-Ligasen, z. B. mit der T4-RNA-Ligase dadurch gezielt an die 5'- oder an die 3'-Seite einer einzelsträngigen Nukleinsäure ligiert werden, indem z. B. ein Ende des Oligonukleotids vor Ligation geschützt wird, während das 5'-Ende phosphoryliert werden kann. Somit kann ein RAT und eine PB enthaltendes Oligonukleotid (Marker und Adapter) gezielt an beide Enden einer RNA ligiert werden. So können RATs z. B. an nicht-proteincodierende RNA ligiert werden und diese im Anschluss daran in DNA umgeschrieben, amplifiziert und sequenziert werden. Entsprechende Konstrukte haben die Domänenabfolge:

5'-PB-RAT-RNA-PB-Amino-C7-3',

oder

5'-PB-RNA-RAT-PB-Amino-C7-3',

wobei alle Moleküle einzelsträngig sind, bzw. PBs auch als doppelsträngige RNA vorliegen können. Alternativ zur Ligation mittels RNA Ligasen kann die RNA mit Hilfe einer Poly(A)-Polymerase künstlich mit einem Poly-A-Strang am 3' Ende der RNA versehen werden.

[0101] Nach Einfügen der Poly-A-Sequenz und Umwandlung in eine doppelsträngige cDNA können auch solche RNAs mit den gleichen Konstrukten wie oben für die mRNAs beschrieben, quantifiziert werden. Da nicht für Proteine codierende RNAs in der Regel keine CAP-Struktur besitzen, kann allerdings nicht schon verfahrenstechnisch zwischen Voll-Längen-RNAs und gebrochenen Molekülen unterschieden werden.

4) Konstrukte für die Quantifizierung von Fragmenten genomischer DNA

4.1.) Konstrukte für die Quantifizierung von Bereichen genomischer DNA: Chromatin-Immuno-Precipitation, Digitales Karyotyping, Methylation-Spezifische Digital Karyotyping, Metagenomics

[0102] Ein Verfahren zur quantitativen Bestimmung von DNA-Proteinbindungen basiert auf einer Vorbereitung der Probe durch Immuno-Präzipitation. Solche Verfahren werden häufig für die Bestimmung der Bindestellen für Transkriptionsfaktoren verwendet. Dabei wird z. B. genomische DNA, die nicht von gebundenen Proteinen befreit ist, enzymatisch mit einem Restriktionsenzym oder einer unspezifisch schneidenden Endonuklease verdaut oder anderweitig (z. B. durch Ultraschall) zerkleinert. Die DNA/Protein Komplexe werden durch Immuno-Präzipitation mit einem Antikörper gegen ein bestimmtes Protein (e. g. einen Transkriptionsfaktor) ausgefällt. Die mitpräzipitierte, an das Protein gebundene DNA wird von allen Proteinen befreit und mit oder ohne Tagging Enzym in eine mittels RATs quantifizierbare Form gebracht. Dazu werden Marker und gegebenenfalls Adapter mit den Enden der präzipitierten DNA-Fragmente ligiert, die enthaltenen Nukleinsäurekonstrukte werden amplifiziert und im Anschluss sequenziert. Als Ergebnis wird die Menge an Kopien einer bestimmten genomischen DNA-Sequenz, z. B. in einer Gewebeprobe, erhalten, an die oder in deren Nähe ein Protein gebunden war.

[0103] Das „Digitale Karyotypisieren (DK) ermöglicht die Quantifizierung von kurzen DNA-Abschnitten im Genom. Die Technik dient dazu, chromosomale Veränderungen, Vervielfältigungen und Deletionen zu untersuchen und die Anwesenheit von Fremd-DNA festzustellen. Hierbei wird die zu untersuchende DNA zunächst mit einem (methylierungsinsensitiven) Restriktionsenzym verdaut. Nun können biotinylierte Marker oder Adapter ligiert werden. Die Probe wird hiernach mit einem zweiten, häufig schneidenden (methylierungsinsensitiven) Restriktionsenzym verdaut und an die entstehenden Schnittstellen Marker oder Adapter ligiert. Die entstehenden Konstrukten müssen mit mindestens einem Marker versehen sein und können so amplifiziert und sequenziert werden.

[0104] Sind Marker oder Adapter mit TES ausgestattet kann ein Tagging Enzym verwendet werden und nach Binden an eine feste Phase entweder die gebundene oder die nicht gebundene Marker-TAG bzw. Adapter-TAG Kombination gereinigt werden und mit einem weiteren Marker oder Adapter versehen werden, so dass je ein Marker mit einem Adapter (oder einem Marker) einen TAG einschließt. Sind Marker und Adapter mit einer TES, insbesondere für EcoP15I versehen wie dies in einer bevorzugten Version vorgesehen ist, so können zudem gebundene und nicht-gebundene Marker-TAG und Adapter-TAG Kombinationen miteinander ligiert werden, so dass zwei TAGs von Marker und/oder Adapter eingeschlossen werden.

[0105] Methylierungsspezifische digitale Karyotypisierung (MSDK): Diese Technik dient dazu, den Methylierungszustand einer genomischen DNA zu untersuchen. Hierzu wird das bei DK verwendete erste Enzym mit einem methylierungssensitiven Enzym ersetzt. Dieses kann nur nicht-methylierte Erkennungssequenzen schneiden. Die restlichen Schritte sind analog zum DK. Je nach Methylierungszustand entstehen so unterschiedliche TAGs aus einer Probe.

[0106] Im Folgenden wird eine besonders bevorzugte Variante des DK beschrieben, die den Einsatz des Tagging-Enzyms EcoP15I einschließt, um Nukleinsäurekonstrukte zur quantitativen Bestimmung von genomischer DNA gemäß dem erfindungsgemäßen Verfahren herzustellen.

[0107] Zur Analyse von genomischen DNA-Fragmenten durch DK, für Metagenomics-Anwendungen bzw. nach Immuno-Präzipitation, wird in einer bevorzugten Methode die DNA wie zuvor beschrieben zunächst mit einem oder mehreren Restriktionsenzymen verdaut. Für die Analyse von aus Immuno-präzipitiertem Material stammender DNA kann diese auch z. B. durch Scherung oder Ultraschall mechanisch zerkleinert werden.

[0108] Die erhaltenen DNA-Fragmente werden hiernach mit einem biotinylierten Oligonukleotid 1 (Adapter oder Marker) ligiert, das distal zum biotinylierten Ende eine Schnittsequenz für ein Restriktionsenzym besitzt, wobei nach dem Restriktionsverdau mit dem entsprechenden Enzym Ligationstelle (LS1) mit einem überhängenden Einzelstrang entsteht. Für eine Ligation mit mechanisch zerkleinerter DNA werden Oligonukleotide (Marker oder Adapter) mit glatten Enden verwendet. Die Enden der zerkleinerten DNA werden vor der Ligation mit dem Klenow Fragment ebenfalls geglättet. Weiterhin enthält der Adapter am biotinylierten Ende

eine Primerbindungsstelle (PB1). Für den Einsatz von EcoP15I als Tagging-Enzym muss das Oligonukleotid 1 zudem die Sequenz 5'-CAGCAG'-3' (TES) vor der Ligationsstelle tragen. Nach Binden an eine Streptavidin-modifizierte feste Phase und Waschen zur Entfernung von ungebundenen Fragmenten wird mit einem weiteren Enzym verdaut. Bevorzugt werden dazu Restriktionsenzyme, die möglichst viele Restriktionsschnittstellen in der zu analysierenden DNA aufweisen, wie z. B. NlaIII, DpnII, TaqI, FatI, Sau3A, MboI etc. verwendet. An die neue Restriktionsschnittstelle wird ein weiteres Oligonukleotid (Oligonukleotid 2) mit an die gebundene DNA ligiert. Das Oligonukleotid 2 enthält eine weitere Erkennungssequenz für das Restriktionsenzym EcoP15I (5'-CTGCTG'-3' (TES)) gefolgt von einer Restriktionsenzymsschnittstelle, aus der nach Restriktionsverdau eine Ligationsstelle (LS2) hervorgeht. Die Ligation erfolgt mit den, durch den Verdau mit dem entsprechenden häufig schneidenden Restriktionsenzym entstehenden Ligationsstellen der DNA. Oligonukleotid 2 ist zudem am distal zur LS befindlichen Ende vor Ligation geschützt, z. B. durch eine Amino-C7 Modifikation.

[0109] Durch die beschriebenen Verfahrensschritte entsteht das folgende Produkt:
5'-CTGCTG-LS2-Nukleinsäure-LS1-CAGCAG-PB1-3'-(feste Phase)

[0110] Nach Verdau mit dem Tagging-Enzym EcoP15I entstehen 25–27 Bp lange TAGs. Diese können nun mit einem weiteren Oligonukleotid 3 ligiert werden. Dieses Oligonukleotid hat den Aufbau RAT-PB2, wobei eine Ligation nur über das freie Ende des RAT erfolgt, da das freie Ende des PB2 vor einer Ligation, z. B. durch eine Amino-C7 Modifikation, geschützt ist. Nach Waschen der festen Phase werden die gebundenen Nukleinsäurekonstrukte von der festen Phase abgelöst, es entsteht folgendes zur Hochdurchsatz-Sequenzierung besonders geeignete Konstrukt:
PB2-RAT-TAG-LS1-CAGCAG-PB1.

[0111] Alternativ oder zusätzlich kann der RAT auch im PB1 enthaltenden Oligonukleotid 1 integriert werden, es entsteht das Produkt PB2-TAG-LS1-CAGCAG-RAT-PB1 bzw. PB2-RAT-TAG-LS1-CAGCAG-RAT-PB1.

[0112] Nach der Herstellung des Zweitstrangs mit einer Polymerase kann durch den Verdau mit einem für die Restriktionsschnittstelle spezifischen Enzym ein Linker hergestellt werden, der die für die Restriktionsschnittstelle typischen Basenüberhänge hat und somit die Ligation mit einer Schnittstelle mit komplementären Basenüberhängen ermöglicht. Alternativ kann die Ligation auch mit Nukleinsäurefragmenten mit glatten Schnittstellen erfolgen. So ist insbesondere die direkte Verknüpfung eines RATs mit einem TAG ohne die Zwischenschaltung eines Linkers möglich. Prinzipiell kann durch Einführung von Restriktionsschnittstellen über eine Linkersequenz, die zur Erzeugung von überhängenden Schnittstellen dienen, der Einbau von Polynukleotiden in die Nukleinsäurekonstrukte in der gewünschten Orientierung erreicht werden.

[0113] Der Abbau kann mechanisch, z. B. mittels Scherkräften, physikalisch, z. B. mittels Ultraschall oder bevorzugt durch enzymatischen Abbau erfolgen.

4.2.) Konstrukte für die Quantifizierung beider Enden von Fragmenten genomischer DNA mit Hilfe einer Zirkularisierungsmethode

[0114] Wie bereits erwähnt, ist es oft gewünscht, von beiden Enden eines Nukleinsäuremoleküls einer genomischen DNA je einen TAG zu gewinnen, um beide Enden zu quantifizieren. Zur Fragmentierung der genomischen DNA kann diese entweder mit Restriktionsenzymen, unspezifischen Endo- oder Exonukleasen oder durch mechanische Zerkleinerung in kleinere Fragmente beliebiger Größe zerlegt werden. Nach Bindung dieser linearen Moleküle an die bereits erwähnten Konstrukte der Form (1) TES-RAT-PB-RAT-TES oder (2) TES-RAT-PB-PB'-RAT-TES können die Moleküle zirkularisiert und danach amplifiziert werden, wenn die Fragmente eine bestimmte Größe nicht überschreiten.

[0115] Diese Reaktionen können an einer festen Phase oder bevorzugt in Lösung stattfinden. Für die Bindung an eine feste Phase bietet sich die Modifikation eines oder beider TES durch eine Biotin-Gruppe an.

[0116] Die Amplifikation des zirkulären Moleküls geht dabei von den Primerbindungsstellen (PB) aus, wobei bei (1) die Primer zumindest teilweise komplementär zueinander und zu je einem Strang der PB sind. Bei (2) handelt es sich um zwei verschiedene Primerbindungsstellen und die von den daran bindenden Primern ausgehende Amplifikation erfolgt gegenläufig. In beiden Fällen entstehen zirkuläre Moleküle, die nach dem Schnitt mit dem TE die Domänen-Abfolge

TAG-TES-RAT-PB-RAT-TAG

bzw.

TAG-TES-RAT-PB-PB'-RAT-TAG

besitzen. Es ist oft sinnvoll (z. B. für eine gerichtete Sequenzierung oder Klonierung), an eines oder beide Enden dieser Konstrukte eine oder zwei weitere Domänen bestehend aus einem oder zwei Adapter(n), ein oder zwei Linker(n) oder einer oder zwei Primer-Bindungsstelle(n) zu binden, so dass dann z. B. Konstrukte mit der in s) bis v) dargestellten Domänen-Folge entstehen:

- cc) PB"-TAG-TES-RAT-PB-RAT-TAG
- dd) PB"-TAG-TES-RAT-PB-RAT-TAG-PB"
- ee) PB"-TAG-TES-RAT-PB-PB'-RAT-TAG-PB" und
- ff) PB"-TAG-TES-RAT-PB-PB'-RAT-TAG-PB'''

Dabei können die Domänen PB" und PB''' hier im Sinne der Erfindung durch die oben erwähnten Domänen (Adapter) ersetzt werden.

Ausführungsbeispiele:

Beispiel 1: Verwendung von RATs für die Quantifizierung von mRNAs.

[0117] In Folgenden wird die Verwendung der RATs für die Quantifizierung von mRNAs in einem bevorzugten Verfahren beispielhaft beschrieben. Dazu wird die mRNA zunächst in eine cDNA umgeschrieben. Dann wird von praktisch jedem cDNA-Molekül ein TAG gewonnen, der mit einem Marker enthaltend einen RAT versehen ist. Das erhaltene Nukleinsäurekonstrukt wird dann amplifiziert und sequenziert.

cDNA-Synthese mit einem oligo-dT Primer:

[0118] Mit Hilfe eines handelsüblichen Kits mit Trizol Reagenz (Invitrogen Inc.) wird etwa 1 mg Gesamt-RNA aus einer biologischen Probe (einer MVV Leukemiezelllinie) gewonnen. Aus der Gesamt-RNA werden mit einem handelsüblichen Kit ("Oligtex-Midi-Kit", Qiagen N. V.) 5 µg mRNA isoliert. Die cDNA wird im Anschluss mit dem „cDNA synthesis system“, (Invitrogen Corp.) hergestellt, wobei das folgende, am 5' Ende biotinyliertes Oligonukleotid verwendet wird, dieses enthält die Erkennungssequenz für EcoP15I CAGCAG:

Biotin-5'-CTGATCTAGAGGTACCGGATCCCAGCAGTTTTTTTTTTTTTTTTTTT-3'

(SEQ ID No. 2)

[0119] Das Produkt wird dann zu einer doppelsträngigen DNA konvertiert (cDNA synthesis system“, Invitrogen Corp.) und in 20 µl LoTE Puffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA) gelöst.

Verdau mit dem Anchoring Enzym DpnII:

[0120] Die doppelsträngige cDNA (20 µl) wird in einer 200 µl Reaktionslösung mit 50 Units DpnII (New England BioLabs Inc, Ipswich, UK; NEB) mitgelieferten Puffer bei 37°C für 90 Minuten verdaut. Nach dem Verdau wird die cDNA zur Entfernung des Enzyms mit TE-equilibriertem Phenol/Chlorophorm/Isoamylalkohol (25:24:1, pH = 8,0) extrahiert, mit Ethanol präzipitiert und in 100 µl LoTE Puffer gelöst. 300 µl Streptavidin-beschichtete magnetische Partikel (MP; Dynabeads M279, Dynal Biotech GmbH, Hamburg, DE) werden mit 200 µl 1 × B&W Lösung (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) gewaschen, wobei nach Zugabe der B&W Lösung der „Magnetic Particel Capturer (MPC; Promega GmbH, Mannheim DE) benutzt wird, um die magnetischen Partikel an der Gefäßwand zu halten und die B&W-Lösung auszutauschen. Zur cDNA (100 µl) werden hiernach 100 µl 2 × B&W Lösung (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) gegeben und die 1 × B&W-Lösung der MP mit der cDNA-Lösung ersetzt. Die MPs werden hiernach zur Bindung der biotinylierten cDNA mit den Streptavidin-MPs unter Rotation für 30 Minuten bei Raumtemperatur inkubiert. Mit Hilfe eines Magneten werden die MPs, an die die DpnII-verdaute cDNA gebunden ist 2 × mit 200 µl B&W-Lösung gewaschen und in 200 µl LowTE rückgelöst.

Adapter-Ligation:

[0121] An die gebundene DNA wird ein Adapter ligiert, der die Erkennungssequenz für das Tagging-Enzym enthält. Dieser ist aus den folgenden einzelsträngigen Oligonukleotiden A und B zusammengesetzt: Oligonukleotid B enthält eine Amino-C7-Modifikation am 3' Ende, um eine ungewollte Ligation an dieser Stelle zu verhindern und ist am 5' Ende phosphoryliert, um die Ligation auf dieser Seite zu ermöglichen. Der mit der Amino-C7-Modifikation geschützte Teil des 5'-Endes des später doppelsträngigen Linkers dient als Primer-Stelle für die spätere Pyrosequenzier-Reaktion mit dem Solexa-Verfahren.

[0122] Dazu werden die beiden Einzelstränge des Oligonukleotid-Adapters

A: 5' -Phosphat-GATCCTGCTGTAGAACTCTGAACCTGT [Amino-C7]-3' (SEQ ID No. 3)

und

B: 5' -ACAGGTTTCAGAGTTCTACAGCAG-3' (SEQ ID No. 4)

in 0,2 mM Tris-EDTA in einer Konzentration von 100 µM gelöst, und miteinander zu einem Doppelstrang hybridisiert, indem beide Oligonukleotid-Lösungen im Verhältnis 1:1 gemischt, auf 96°C erhitzt und danach langsam auf Raumtemperatur abgekühlt werden. Der Linker enthält die Erkennungssequenz für das Tagging-Enzym EcoP15I (CAGCAG). Es entsteht ein doppelstängiger Linker mit einem 5' GATC-Überhang, der mit der DpnII Restriktionsschnittstelle ligiert werden kann.

Ligase-Reaktion:

[0123] Folgender Ligationsansatz wird hergestellt:

6 µl 5 × Ligase Buffer (Invitrogen, Corp.)

+5 µl Linker (100 µM)

+17 µl H₂O

= 28 µl

[0124] Das LoTE des MP + cDNA-Ansatzes wird durch den 28 µl Ligations-Ansatz ersetzt. Die MP werden vorsichtig gemischt, für 2 Minuten auf 50°C erhitzt und 15 Minuten bei Raumtemperatur abgekühlt. 2 µl T4-Ligase (5 U/µl, Invitrogen, Corp.) werden zu dem Ligationsansatz gegeben und die Probe bei 16°C für 2 Stunden inkubiert, alle 20 Minuten wird die Probe vorsichtig gemischt. Nach der Ligationsreaktion werden die nicht-gebundenen Linker von den MP durch 3-maliges waschen mit 300 µl 1 × B&W entfernt. Die MPs werden in ein neues Gefäß überführt und erneut 1 × mit 300 µl 1 × B&W und anschließend 2 × mit 1 × NEB4-Puffer (New England Biolabs, Inc., NEB) gewaschen und in 300 µl NEB4-Puffer aufgenommen.

Schneiden mit dem Tagging-Enzym:

[0125] Folgender Reaktionsansatz wird vorbereitet:

10 µl 10 × Puffer NEB3

10 µl 10 mM ATP (NEB)

2 µl EcoP15I (NEB)

1 µl 100 × BSA (NEB) 77 µl H₂O

= 100 µl

[0126] Die 300 µl NEB3-Puffer des MP + cDNA + Linker-Ansatzes wird durch den 100 µl Reaktionsansatz ersetzt. Die Probe wird bei 37°C eine Stunde bei mehrmaligem vorsichtigem Mischen inkubiert. Mithilfe des MPC werden die MP an der Gefäßwand festgehalten, während der Überstand abgenommen und in ein neues Gefäß überführt werden kann. Der Überstand enthält die Linker-TAGs von 52 bp Länge (Linker + TAG). Die Linker-TAG-Fragmente im 100 µl Überstand werden im Verhältnis 1 zu 1 mit Phenol/Chlorophorm/Isoamylalkohol (25:24:1, pH = 8,0) Lösung extrahiert, mit 500 µl Ethanol + 67 µl 7.5 M NH₄OAc + 2 µl Glycogen (NEB) präzipitiert. Nach Inkubation bei -20°C für 8 h wird die Probe 30 Minuten bei 4°C und 13.000 g zentrifugiert, zwei mal mit 70%-igem Ethanol gewaschen, getrocknet und in 20 µl LoTE Puffer gelöst.

Herstellung glatt-endender „blunting“ der Adapter-TAG Fragmente:

[0127] Durch Schnitt mit EcoP15I entstehen Adapter-TAG-Fragmente mit überstehenden 5'-Enden. Um daran später einen RAT-enthaltenden Marker mit glatten Enden anzuliegender werden die Enden geglättet (blunting). Die blunting-Reaktion findet in 50 µl 1 × NEBuffer 2 (New England Biolabs, Inc.) mit 33 µM dNTP, 5 units Polymerase I, Large (Klenow) Fragment (New England Biolabs Inc.), bei 25°C und mit einer 15 minütigen Inkubationszeit statt. Die Probe wird hiernach mit LoTE auf 200 µl aufgefüllt und mit Phenol/Chlorophorm/

Isoamylalkohol (25:24:1, pH = 8,0) Lösung extrahiert, mit 1000 µl Ethanol + 133 µl 7.5 M NH₄OAc + 2 µl Glycogen (NEB) präzipitiert und in 10 µl LoTE gelöst.

Herstellung und Ligation mit dem RAT-enthaltenden Marker:

[0128] Der einzelsträngige Marker mit folgender Sequenz:

Phosphat 5' NNNNNNNNTCGTATGCCGTCTTCTGCTTG 3'Amino-C7

(=RAT-A), wobei (N)₈den einzelstängigen RAT darstellt (SEQ ID No. 5)

wird mit dem Primer mit folgender Sequenz

[0129]

5' CAAGCAGAAGACGGCATACGA-3' (= RAT-B) (SEQ ID No. 6)

mithilfe des Klenow Fragments in folgender Lösung für 30 Minuten bei 37°C, inkubiert, und somit doppelsträngig hergestellt:

5 µl RAT A (10 µM)

5 µl RAT B (10 µM)

1 µl dNTPs (10 mM each)

5 µl 10 × NEBuffer 2 (New England Biolabs Inc.)

1 µl Klenow Fragment (New England Biolabs Inc.)

+ 33 µl H₂O

= 50 µl

[0130] Der doppelsträngige, den RAT-enthaltende Marker wird mittels Phenol-Chlorophorm gereinigt, mit Ethanol präzipitiert gewaschen und nach Trocknen in 50 µl H₂O aufgenommen.

[0131] Die Ligation mit dem Adapter-TAG-Konstrukt findet unter folgenden Bedingungen statt:

10 µl TAG-Lösung in LoTE

1 µl 0,1 mM Marker

4 µl 5 × Ligation Buffer (Invitrogen, Inc.)

2 µl T4-Ligase (Invitrogen, Inc.)

3 µl H₂O

= 20 µl

[0132] Der Ansatz wird bei 16°C für 2 Stunden inkubiert. Die Probe wird hiernach auf 200 µl mit LoTE aufgefüllt und mit Phenol/Chlorophorm/Isoamylalkohol (25:24:1, pH = 8,0) Lösung extrahiert, mit 1000 µl Ethanol + 133 µl 7.5 M NH₄OAc + 2 µl Glycogen (NEB) präzipitiert und in 10 µl LoTE gelöst.

PCR-Amplifikation:

[0133] Zunächst wird mit verschiedenen Verdünnungen der Linker-TAG-RAT-Ligationsansatz eine PCR mit 25 Zyklen durchgeführt, um die am besten für die Amplifikation geeignete Konzentration zu ermitteln. Die PCR-Reaktionsansätze zu je 20 µl werden für zur initialen Denaturierung der Templat-DNA für 2 Minuten bei 98°C inkubiert, es folgen 25 Zyklen mit 98°C für 15 Sekunden, gefolgt von 15 Sekunden bei 60°C und abschließend 15 Sekunden bei 72°C. Als Polymerase wird die Phusion-Polymerase (Finnzymes, Finnland) benutzt.

Als Primer wurden verwendet

5' CAAGCAGAAGACGGCATACGA 3' (SEQ ID No. 7)

Und

5' ACAGGTTTCAGAGTTCTACAGCAG 3' (SEQ ID No. 8)

[0134] Die PCR-Produkte werden auf ein 8% iges Polyacrylamidgel aufgetragen. Nach Färben mit Ethidiumbromid und Visualisieren unter UV-Licht wird eine 79 Bp Marker-Adapter-TAG Bande im Gel sichtbar. Danach

wird für präparative Zwecke ein 200 µl PCR-Ansatz mit derjenigen Verdünnung durchgeführt, welche die beste Amplifikation ergab.

Reinigung der Linker-TAG-RAT Fragmente:

[0135] Die Produkte der präparativen PCR werden erneut auf ein 8% iges Polyacrylamidgel aufgetragen und das Gelstück mit der 79 Basenpaar (Bp) Bande mit einem Skalpell herausgeschnitten. In einem 1,5 ml Eppendorf-Gefäß wird das Gelstück mit einem Pistill zerkleinert und in 500 µl einer Elutionslösung (0,5 % NH₄OAc, 2 mM EDTA 0,1% SDS) über Nacht bei 37°C inkubiert. Das eluierte RAT-Linker-TAG-Konstrukt wird anschließend mit Ethanol aus der Elutionslösung gefällt, 2 × mit 70% Ethanol gewaschen, getrocknet und in 10 µl LoTE rückgelöst.

[0136] Die Konzentration und Qualität des PCR Produktes wird mit einer Gelelektrophorese bestimmt. Zum Test wird ein Aliquot des gereinigten RAT-Linker-TAG-Konstrukts in den pGEM®-T Easy Vektor (Promega) ligiert, in den E. coli-Stamm DH5alpha transformiert, weiße-Kolonien selektiert und 5–10 RAT-Linker-TAG-Konstrukte enthaltende Plasmide mit dem Sanger-Verfahren sequenziert. Sieben der dabei erhaltenen, typischen Sequenzen mit unterschiedlichen PB-RAT-TAG-LS-TES-PB'-Konstrukten sind unten aufgelistet. Jeder TAG ist eindeutig mit einem RAT markiert. Die RAT-Domänen der Konstrukte sind dabei fett markiert, die TAG-Domänen kursiv dargestellt.

Klon 1

CAAGCAGAAGACGGCATACGAT**CTCGACG**GCCACATCCTCGGTATGACAGGAG
ATCCTGCTGTAGAACTCTGAA (SEQ ID No. 9)

Klon 2

CAAGCAGAAGACGGCATACGAG**TCAAGCG**TGGGGGTGGGGGAGACACAATTT
GATCCTGCTGTAGAACTCTGAA (SEQ ID No. 10)

Klon 3

CAAGCAGAAGACGGCATACGA**ACTTACA**ATTCTATGGTAGTACACACACACAG
GATCCTGCTGTAGAACTCTGAA (SEQ ID No. 11)

Klon 4

CAAGCAGAAGACGGCATACGAT**TGATACC**TTCTCCAGGGCTCTGGTTACCTGA
TCCTGCTGTAGAACTCTGAA (SEQ ID No. 12)

Klon 5

CAAGCAGAAGACGGCATACGAC**CCCATCCCC**CAAGGCTGATATTTCCCTGGGA
TCCTGCTGTAGAACTCTGAA (SEQ ID No. 13)

Klon 6

CAAGCAGAAGACGGCATACGAT**GTATTTG**ATTTATAATTTAATGGCTTGCAGATC
CTGCTGTAGAACTCTGAACCTGT (SEQ ID No. 14)

Klon 7

CAAGCAGAAGACGGCATACGAT**GATGTATCCAATGGTTCCAGTCACATATTGAT**
 CCTGCTGTAGAACTCTGAACCTGT (SEQ ID No. 15)

[0137] Mit Hilfe einer BLAST-Recherche mit dem RAT-TAG Konstrukt in GenBank konnten die einzelnen TAGs jeweils einem bestimmten humanen Gen zugeordnet werden (Tabelle 1):

Tabelle 1: Tabellarische Zusammenstellung der Ergebnisse der BLAST-Recherche

Klon Nr.	Ergebnis der BLAST-Recherche
Klon 1	>gi 14993673 gb AC010366.6 Homo sapiens chromosome 5 clone CTD-2044J15, complete sequence Length=115396 Score = 54.0 bits (27), Expect = 2e-05 Identities = 27/27 (100%), Gaps = 0/27 (0%) Strand=Plus/Plus Query 9 GCCACATCCTCGGTATGACAGGAGATC 35 Sbjct 6328 GCCACATCCTCGGTATGACAGGAGATC 6354

Klon 2	>gi 23272853 gb BC035646.1 Homo sapiens hypothetical protein LOC284356, mRNA (cDNA clone IMAGE:5558805), with apparent retained intron Length=1330 Score = 56.0 bits (28), Expect = 5e-06 Identities = 28/28 (100%), Gaps = 0/28 (0%) Strand=Plus/Minus Query 8 GTGGGGGTGGGGGAGACACAATTTGATC 35 Sbjct 1285 GTGGGGGTGGGGGAGACACAATTTGATC 1258
Klon 3	>gi 10716633 gb AC009404.5 Homo sapiens BAC clone RP11-28H22 from 2, complete sequence Length=112883 Score = 58.0 bits (29), Expect = 1e-06 Identities = 29/29 (100%), Gaps = 0/29 (0%) Strand=Plus/Plus Query 8 ATTCTATGGTAGTACACACACACAGGATC 36 Sbjct 98244 ATTCTATGGTAGTACACACACACAGGATC 98272
Klon 4	>gi 142381010 ref NM_002488.3 Homo sapiens NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa (NDUFA2), mRNA Length=599 Score = 52.0 bits (26), Expect = 8e-05 Identities = 26/26 (100%), Gaps = 0/26 (0%) Strand=Plus/Minus Query 9 TTCTCCAGGGCTCTGGTTACCTGATC 34 Sbjct 350 TTCTCCAGGGCTCTGGTTACCTGATC 325
Klon 5	>gi 146147624 gb BC106894.2 Homo sapiens cDNA clone IMAGE:40030896 Length=673 Score = 52.0 bits (26), Expect = 8e-05 Identities = 26/26 (100%), Gaps = 0/26 (0%) Strand=Plus/Plus Query 9 CCCAAGGCTGATATTTCCCTGGGATC 34 Sbjct 48 CCCAAGGCTGATATTTCCCTGGGATC 73

Klon 6	>gi 18376860 gb AC011345.5 Homo sapiens chromosome 5 clone CTC-293E5, complete sequence Length=105797 Score = 44.1 bits (22), Expect = 0.018 Identities = 22/22 (100%), Gaps = 0/22 (0%) Strand=Plus/Plus Query 10 GCCATTAAATTATAAATCAAAT 31 Sbjct 27015 GCCATTAAATTATAAATCAAAT 27036
Klon 7	>gi 164692865 dbj AK313066.1 Homo sapiens cDNA, FLJ93545 Length=1860 Score = 52.0 bits (26), Expect = 8e-05 Identities = 26/26 (100%), Gaps = 0/26 (0%) Strand=Plus/Plus Query 1 GATCAATATGTGACTGGAACCATTGG 26 Sbjct 1555 GATCAATATGTGACTGGAACCATTGG 1580

Es folgt ein Sequenzprotokoll nach WIPO St. 25. Dieses kann von der amtlichen Veröffentlichungsplattform des DPMA heruntergeladen werden.

Patentansprüche

1. Verfahren zur quantitativen Bestimmung von Nukleinsäuren in einer Probe **dadurch gekennzeichnet**, dass

a) Nukleinsäurekonstrukte, enthaltend

– mindestens eine kennzeichnende Polynukleotiddomäne einer, in einer Probe vorhandenen Nukleinsäure, die quantitativ bestimmt werden soll und

– mindestens eine artifizielle Polynukleotiddomäne mit zufälliger Sequenz und

– mindestens eine Primer- oder Polymerasebindungsstelle,

wobei jeweils eine kennzeichnende Polynukleotiddomäne und mindestens eine Polynukleotiddomäne mit zufälliger Sequenz eine charakteristische Domänenkombination bilden die von mindestens einer Primer- oder Polymerasebindungsstelle flankiert wird,

in Anwesenheit einer Polymerase amplifiziert werden und

b) nach der Amplifikation zumindest die Sequenz der Polynukleotiddomänen mit zufälliger Sequenz und eines Teils der kennzeichnenden Polynukleotiddomäne ermittelt werden, wobei zur quantitativen Bestimmung der in der Probe vorhandenen Nukleinsäure sequenzidentische oder sequenzähnliche Kombinationen einer kennzeichnenden Polynukleotiddomäne und einer Polynukleotiddomäne mit zufälliger Sequenz als Kopien einer kennzeichnenden Polynukleotiddomäne der ursprünglich in der Probe vorhandenen Nukleinsäure identifiziert werden können.

2. Verfahren zur quantitativen Bestimmung von Nukleinsäuren nach Anspruch 1, **dadurch gekennzeichnet**, dass es sich bei der Amplifizierung um eine PCR-Amplifizierung oder um eine Amplifizierung durch Klonierung handelt.

3. Verfahren zur quantitativen Bestimmung von Nukleinsäuren nach einem der Ansprüche 1 oder 2 **dadurch gekennzeichnet**, dass

a) eine Probe enthaltend ein Polynukleotid mit einer kennzeichnenden Polynukleotiddomäne mit einer Lösung aus einem Polynukleotid enthaltend eine artifizielle Polynukleotiddomäne mit zufälliger Sequenz versetzt wird, wobei zumindest eine der beiden Polynukleotide eine Primerbindungsstelle besitzt,

- b) eine Ligation mindestens eines Endes der Polynukleotide enthaltend die kennzeichnenden Polynukleotiddomänen mit dem Polynukleotid enthaltend die Polynukleotiddomäne mit zufälliger Sequenz, so dass die Primerbindungsstelle nicht zwischen der kennzeichnenden Polynukleotiddomäne und der Polynukleotiddomäne mit zufälliger Sequenz liegt und
- c) die erhaltenen, mit einer Polynukleotiddomäne mit zufälliger Sequenz markierten Nukleinsäurekonstrukte durch Amplifikation mit einer Polymerase vervielfältigt und zumindest die zufällige Sequenz und die Sequenz der kennzeichnenden Polynukleotiddomäne ermittelt wird.

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet**, dass ein Ende der Polynukleotide enthaltend eine kennzeichnende Polynukleotiddomäne und/oder ein Ende der Polynukleotide enthaltend eine Polynukleotiddomäne mit zufälliger Sequenz vor Ligation und/oder enzymatischem Verdau geschützt werden, wobei, sofern eine Primerbindungsstelle in dem Polynukleotid enthalten ist, dieses Ende vor Ligation und/oder enzymatischen Abbau geschützt ist.

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet**, dass die an einem Ende vor Restriktionsverdau geschützten Polynukleotide, die eine kennzeichnende Polynukleotiddomäne enthalten, mit einem Restriktionsenzym verdaut werden bevor die Polynukleotide enthaltend eine Polynukleotiddomäne mit zufälliger Sequenz an das freie Ende des verbleibenden geschützten Polynukleotidfragmentes ligiert wird.

6. Verfahren nach einem der Ansprüche 3 bis 5, **dadurch gekennzeichnet**, dass die Polynukleotide eine Bindungsstelle zur Anbindung eines Tagging-Enzyms enthalten und nach Ligation der Polynukleotide enthaltend die kennzeichnenden Polynukleotiddomänen mit den Polynukleotiden enthaltend eine Polynukleotiddomäne mit zufälliger Sequenz und nach Anbindung des Tagging-Enzyms aus den erhaltenen ligierten Polynukleotiden durch das Tagging-Enzym Nukleinsäurekonstrukte abgespalten werden, die eine kennzeichnende Polynukleotiddomäne und die eine Polynukleotiddomäne mit zufälliger Sequenz enthalten.

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet**, dass das eingesetzte Tagging-Enzym ein Typ II oder Typ III-Restriktionsenzym ist.

8. Verfahren nach einem der Ansprüche 6 oder 7, **dadurch gekennzeichnet**, dass das verwendete Tagging-Enzym EcoP15I, MmeI oder BsmFI ist.

9. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, dass das verwendete Nukleinsäurekonstrukt eine der folgenden Domänenabfolgen aufweist:

PB-TAG-RAT,
 PB-RAT-TAG,
 PB-TAG-RAT-PB', PB-RAT-TAG-RAT,
 PB-RAT-TAG-RAT-PB',
 PB-RAT-TAG-TAG-RAT,
 PB-RAT(TAG)_n-PB; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10, PB-RAT-(TAG)_n-RAT-PB; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10,
 PB-RAT-(TAG)_n; mit n = 2 bis 50, bevorzugt mit n = 2 bis 10,
 PB-RAT-TAG-RAT-TAG,
 PB-TAG-RAT-TAG-RAT,
 PB-RAT-TAG-RAT-TAG-PB',
 PB-(RAT-TAG)_n-PB'; mit n = 3 bis 50, bevorzugt mit n = 3 bis 10,
 PB-RAT-TAG-PB'-PB''-RAT-TAG-PB''',
 PB-RAT-TAG-PB'-PB''-RAT-TAG-PB'''-PB''''-RAT-TAG-PB''''',
 PB-RAT-TAG-PB'-(PB-RAT-TAG-PB')_n-PB-TAG-RAT-PB';
 mit n = 0 bis 10,
 PB-TES-TAG-RAT,
 PB-TAG-TES-RAT,
 PB-RAT-TAG-TES,
 PB-RAT-TES-TAG,
 PB-TAG-TES-RAT-PB',
 PB-TES-TAG-RAT-PB',
 PB-RAT-TAG-TES-RAT,
 PB-RAT-TES-TAG-RAT,
 PB-RAT-TAG-TES-RAT-PB',
 PB-RAT-TES-TAG-RAT-PB',

PB-RAT-(TES)_m-TAG-TAG-(TES)_m-RAT;

mit $m = 0$ oder 1 unabhängig voneinander für jedes m , wobei mindestens ein $m = 1$ ist,

PB-RAT-TES-TAG-RAT-TAG,

PB-RAT-TAG-TES-RAT-TAG,

PB-TES-TAG-RAT-TAG-RAT,

PB-TAG-TES-RAT-TAG-RAT,

PB-TAG-RAT-TES-TAG-RAT,

PB-TAG-RAT-TAG-TES-RAT,

PB-RAT-TES-TAG-RAT-TAG-PB',

PB-RAT-TAG-TES-RAT-TAG-PB',

PB-RAT-TAG-RAT-TES-TAG-PB',

PB-RAT-TAG-RAT-TAG-TES-PB',

PB-(RAT-(TES)_m-TAG)_n-PB';

mit $n = 3$ bis 50 , bevorzugt mit $n = 3$ bis 10 und $m = 0$ oder 1 unabhängig voneinander für jedes m ,

PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''; mit $m = 0$ oder 1 für jedes m unabhängig voneinander, wobei mindestens ein $m = 1$ ist,

PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''-PB''''-RAT-(TES)_m-TAG-(TES)_m-PB''''';

mit $m = 0$ oder 1 für jedes m unabhängig voneinander, wobei mindestens ein $m = 1$ ist,

PB-RAT-(TES)_m-TAG-(TES)_m-PB'-(PB-RAT-(TES)_m-TAG-(TES)_m-PB')_n-PB-(TES)_m-TAG-(TES)_m-RAT-PB'

mit $n = 0$ bis 10 und mit $m = 0$ oder 1 für jedes m unabhängig voneinander, wobei mindestens ein $m = 1$ ist,

TAG-TES-RAT-PB-RAT-TES,

TAG-TES-RAT-PB-PB'-RAT-TES,

PB''-TAG-TES-RAT-PB-RAT-TAG,

PB''-TAG-TES-RAT-PB-RAT-TAG-PB',

PB''-TAG-TES-RAT-PB-PB'-RAT-TAG-PB'''

TAG-RAT-PB-RAT,

TAG-RAT-PB-PB'-RAT,

PB''-TAG-RAT-PB-RAT-TAG,

PB''-TAG-RAT-PB-RAT-TAG-PB'

PB''-TAG-RAT-PB-PB'-RAT-TAG-PB'''

PB-TAG-RAT-TAG oder

PB-TAG-RAT-TAG-PB', wobei

PB, PB', PB'', PB''', PB''''', PB'''''' Primerbindungsstellen sind die gleiche oder unterschiedliche Sequenzen besitzen können, RAT für eine Polynukleotiddomäne mit zufälliger Sequenz steht, TAG für eine Polynukleotiddomäne mit kennzeichnender Sequenz steht und TES für eine Tagging-Enzym-Bindungsstelle steht.

10. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, dass das verwendete Nukleinsäurekonstrukt eine der folgenden Domänenabfolgen aufweist:

cDNA-Oligo-A/T-RAT-Promotor",

cDNA-Oligo-A/T-RAT-PB-Promotor",

cDNA-Oligo-A/T-TES-RAT-PB,

5'-Promotor-RAT-mRNA-3',

5'-Biotin-PB/Promotor-RAT-TES-mRNA,

5'-Biotin-PB/Promotor-RAT-TES-cDNA-Oligo-A/T-3',

5'-Biotin-PB/Promotor-RAT-TES-TAG-3',

5'-Biotin-PB/Promotor-RAT-TAG,

5'-Biotin-PB/Promotor-TAG-TES-PB-3',

5'-Biotin-PB/Promotor-TAG-TES-RAT-PB-3',

5'-PB-cDNA-Oligo-A/T-PB'-3'

5'-PB-RAT-cDNA-3'

5'-cDNA-Oligo-A/T-PB'-3',

5'-cDNA-Oligo-A/T-RAT'-PB'-3'

5'-TAG-RAT-PB-PB'-RAT'-TAG-3'

5'-PB''-TAG-RAT-PB-PB'-RAT- TAG-PB'''-3',

5'-PB''-TAG-RAT-PB-PB'-TAG-PB'''-3',

5'-TES-cDNA-3'

5'-PB-TES-cDNA-3'

5'-PB-RAT-RNA-PB-Amino-C7-3',

5'-PB-RNA-RAT-PB-Amino-C7-3',

5'-PB-RAT-RNA-PB-Amino-C7-3',

5'-PB-RNA-RAT-PB-Amino-C7-3',
TAG-TES-RAT-PB-RAT-TAG, oder
TAG-TES-RAT-PB-PB'-RAT-TAG.

11. Markerensemble zur Durchführung des Verfahrens gemäß einem der Ansprüche 1 bis 10 zur quantitativen Bestimmung von Nukleinsäuren, **dadurch gekennzeichnet**, dass die enthaltenen Marker jeweils eine artifizielle Polynukleotiddomäne mit voneinander unterschiedlicher zufälliger Sequenz und eine Primerbindungsstelle umfassen.

12. Markerensemble nach Anspruch 11, **dadurch gekennzeichnet**, dass die Marker eine Bindungsstelle für ein Tagging-Enzym enthalten, wobei die Tagging-Enzym-Bindungsstelle am distalen Ende zur Primerbindungsstelle des Markers liegt.

13. Verwendung eines Ensembles aus Markern nach einem der Ansprüche 11 oder 12 zur quantitativen Bestimmung von Nukleinsäuren in einer Probe.

14. Verwendung eines Verfahrens nach einem der Ansprüche 1 bis 8 zur quantitativen Bestimmung von mRNA, cDNA, Aptamer-DNA oder genomischer DNA, zur methylierungsspezifischen quantitativen Bestimmung von genomischer DNA, zur quantitativen Bestimmung von proteingebundener genomischer DNA.

Es folgen 4 Seiten Zeichnungen

1) AGCTGTCAGCCT CTGATGGTGTTGCTAGGTACTGAATT

RAT TAG

2) CATGCTAGTCGG CTGATGGTGTTGCTAGGTACTGAATT

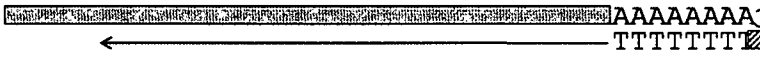
RAT TAG

3) GGCTTATCGTAG CTGATGGTGTTGCTAGGTACTGAATT

RAT TAG

4) GGCTTATCGTAG CTGATGGTGTTGCTAGGTACTGAATT

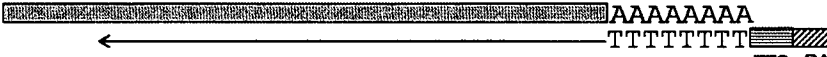
RAT TAG

mRNA 5'  3' **RAT PB PB'**

1.1.1) Alternative: Reverse Transkription mit polyT-TES-RAT-PB-T7

mRNA 5'  3' **TES RAT PB PB'**

1.2) Zweitstrangsynthese

DNA 5'  3' **RAT PB PB'**

1.3) Lineare Amplifikation der Antisense RNA

5'  3' **RAT PB PB'**

RAT PB PB'

RAT PB PB'

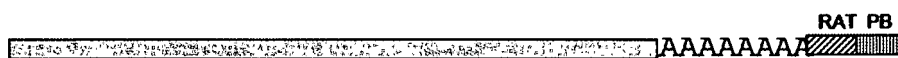
RAT PB PB'

Abb. 2 (Fortsetzung)

1.4) Erststrangsynthese: Reverse Transkription mit Oligohexameren



1.5) Amplifizierte antisense DNA



1.6) Zweitstrangsynthese unter Verwendung eines (wahlweise) biotinylierten Primers komplementär zu PB

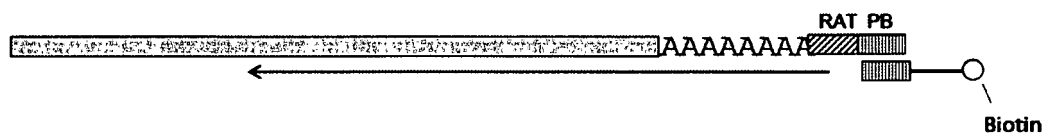
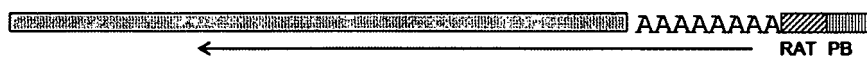


Abb. 2 (Fortsetzung)

2) Analyse von mRNA ohne vorherige Amplifikation



3) Analyse von mRNA ohne vorherige Amplifikation unter Einbau einer TES



4) Amplifikation von mRNA unter Einbau von Markern an einem oder beiden Enden

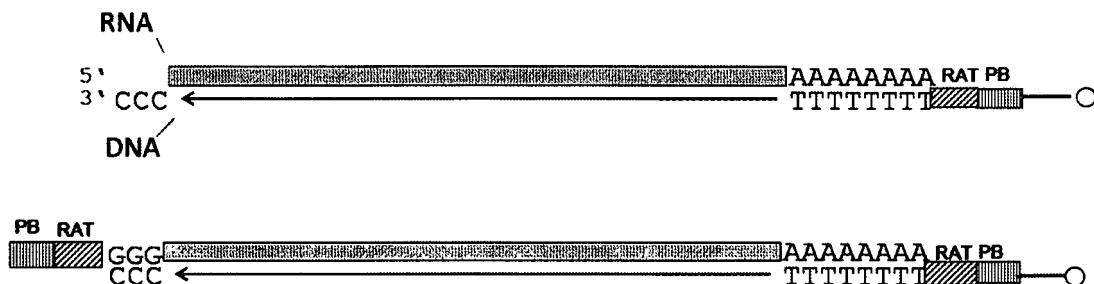
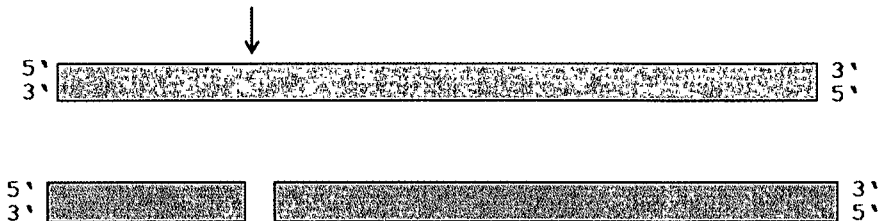


Abb. 3

1. Verdau mit Methylierungsinsensitivem Restriktionsenzym



2. Ligation einer ersten Markers

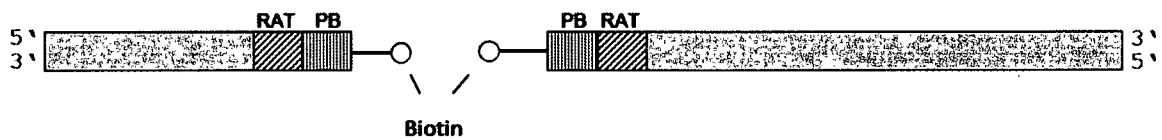
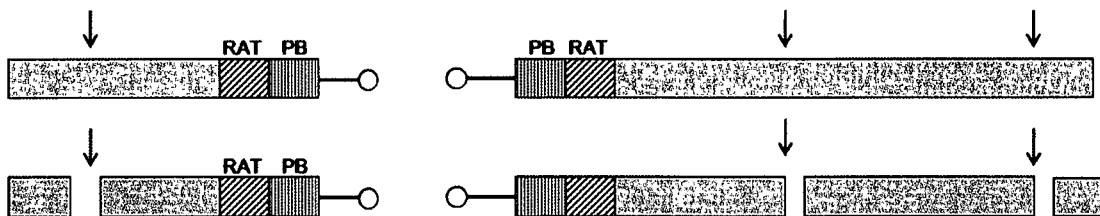
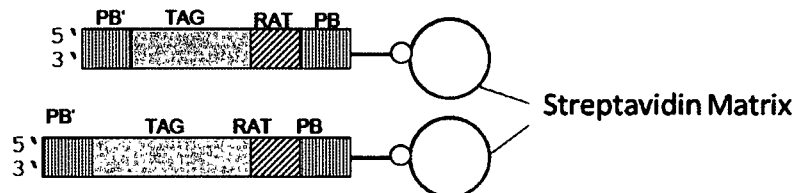


Abb. 3 (Fortsetzung)

3. Verdau mit zweitem Enzym



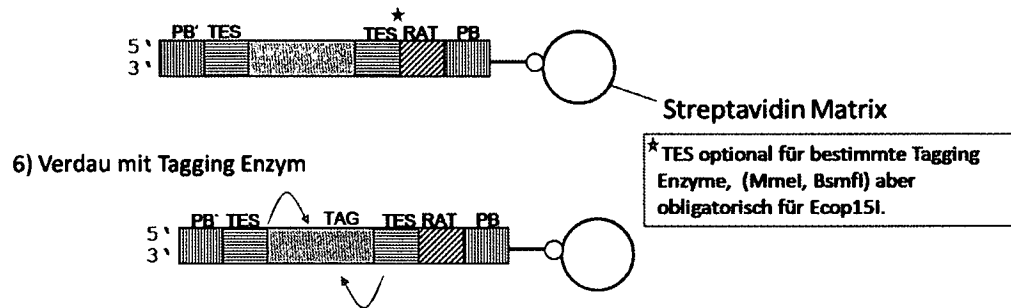
4. Binden an Streptavidin Matrix, Ligation mit Adapter enthaltend PB'



5. Sequenzierung

Abb. 4

5) Nukleinsäurekonstrukte, aufgebaut mit Marker und Adapter, die distal zur PB bzw. PB' eine Tagging Enzym Bindestelle (TES) enthalten ergibt sich an Stelle von Schritt 4) aus Abb. 3



7.1) Verwendung des nicht gebundenen Endes als Tag :



7.2) Verwendung des gebundenen Endes als Tag :



Abb. 4 (Fortsetzung)

7.3) Verwendung beider Enden als Tag :



8.) Sequenzierung