

(10) **DE 10 2008 025 656 B4** 07/28/2016

(12)

Patent specification

(21) File reference number: **10 2008 025 656.0**

(51) Int. class. **C12Q 1/68** (2006.01)

(22) Date of application: **05/28/2008**

(43) Disclosure date: **12/03/2009**

(45) Date of publication of the patent
grant: **07/28/2016**

Within nine months of publication of the patent grant, an opposition may be filed against the patent pursuant to Section 59 of the Patent Act. The objection must be made in writing and must state the reasons on which it is based. An objection fee of €200 must be paid within the objection period (Section 6 of the Patent Costs Act in conjunction with the Annex to Section 2(1) of the Patent Costs Act).

(73) Patent holder:
GenXPro GmbH, 60438 Frankfurt, DE

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

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

(56) State of the art determined:
see following pages

(54) Designation: **Method for the quantitative analysis of nucleic acids, markers therefor and their use**

(57) Main claim: Method for the quantitative determination of nucleic acids in a sample, characterized in that

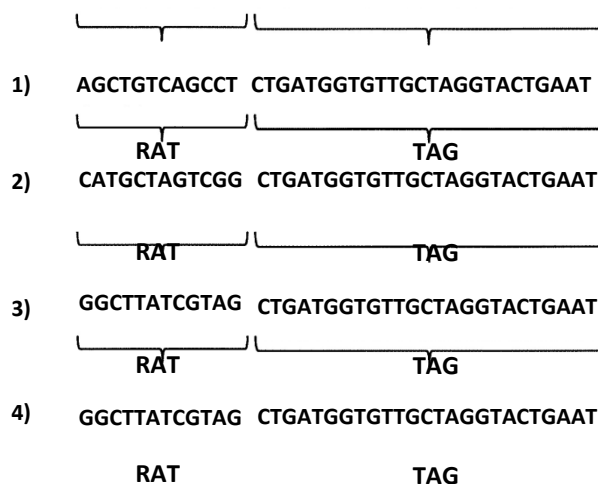
a) nucleic acid constructs containing

- at least one characterizing polynucleotide domain of a nucleic acid present in a sample which is to be quantitatively determined and

- at least one artificial polynucleotide domain with a random sequence and

- at least one primer or polymerase binding site, wherein in each case a characterizing polynucleotide domain and at least one polynucleotide domain with random sequence form a characteristic domain combination which is flanked by at least one primer or polymerase binding site, are amplified in the presence of a polymerase and

b) after amplification, at least the sequence of the polynucleotide domains with random sequence and a part of the characterizing polynucleotide domain are determined, wherein, for the quantitative determination of the nucleic acid present in the sample, sequence-identical or sequence-identical and sequence-like combinations of a characterizing polynucleotide domain and a polynucleotide domain with random sequence can be identified as copies of a characterizing polynucleotide domain of the nucleic acid originally present in the sample.



(56) Determined state of the art:

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

GOWDA, M. [et al.]: 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. [et al.]: SuperSAGE. Cellular Microbiology (2005) 7 (11) 11-18

MEYER, M. [et al.]: Parallel tagged sequencing on the 454 platform. Nat. Protoc. (2008) 3 (2) 267-78 (published electronically on January 31, 2008)

WONG, K.K. [et al.]: 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

Specification

[0001] The invention relates to a method for the quantitative determination of nucleic acids in a sample, in particular for the quantitative determination of gene transcripts, genomic DNA or genomic nucleic acid segments, such as DNA, RNA, mRNA, cDNA, microRNA, non-codingRNA, aptamer DNA and RNA, and to the provision of markers for carrying out the analytical methods.

[0002] The quantification of nucleic acids is of great scientific and commercial importance in all areas of life sciences. For example, it is important to know which gene products (e.g. protein-coding and non-coding RNA) occur how frequently (gene expression studies) and how transcription is regulated, e.g. by epigenetic properties of the genomic DNA. In addition, determining the frequency of copies of gene segments for karyotyping is an important diagnostic option for diseases such as cancer.

[0003] This need has led to the development of various methods, which are generally based either on the non-specific measurement of the total quantity of all or certain nucleic acids by photometric methods (total RNA or DNA), for example, or on the individual quantification of a large number of different nucleic acid molecules of certain sequences. The methods used for this purpose can basically be divided into two groups: The first group consists of methods that can quantify sequences that are already known or characterized in other ways. They are usually based on the binding of the nucleic acids to be quantified to known or otherwise characterized sequences that are bound to a solid surface. Examples include all forms of Southern blots, Northern blots, and so-called DNA or RNA microarrays (chips).

[0004] The second group consists of methods based on quantifying nucleic acids by sequencing as many individual nucleic acid molecules as possible and counting the sequenced molecules in the mixture. Since there are usually more nucleic acid molecules in the mixture than are sequenced, this preserves the relative frequency of a nucleic acid of a specific sequence in the mixture.

[0005] The invention presented here improves the safety and accuracy of nucleic acid quantification by virtually all sequencing and counting-based methods. This includes both so-called “tag”-based methods, in which the tag is a section that is as representative as possible of the nucleic acid and can be used to identify the nucleic acid, and methods in which the number of complete DNA and RNA molecules, such as those of viruses, is determined.

[0006] Sequence-based quantitative methods are becoming increasingly widespread due to the development of techniques for simultaneously sequencing hundreds of thousands to millions of nucleic acid molecules, replacing qualitative, array-based methods in many areas. That is why sequence-based quantification will become increasingly important in the future. A list of the most important sequence-based quantification methods for nucleic acids can be found in S. M. Wang (2007) Understanding SAGE data, Trends in Genetics 23: 42–50. Specifically, this includes:

- “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 facilitates the global, quantitative characterization of a transcriptome by using a type II restriction enzyme (BsmFI) as a “tagging enzyme” to excise a 14-bp fragment, the so-called “tag,” from a cDNA behind the 3'-end interface for a frequently cutting restriction enzyme (usually behind the NlaIII interface CATG). In SAGE, two tags are joined head-to-head to form so-called ditags, which are then linked (“concatenated”) to form longer chains (“concatenamers”). These concatemers are then cloned and sequenced. The tags represent the original mRNA molecules from which the cDNA was generated. Therefore, the number of sequenced tags represents a measure of the relative frequency with which the mRNA was present in the mRNA mixture. Today, there is no need for concatenation and cloning because the dits can be sequenced directly after PCR using high-throughput sequencing techniques such as those described above (e.g., 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) was developed to increase the specificity of the SAGE tags, which are only 14 bp long, by using a different type II restriction enzyme (MmeI) to generate tags that are 21 bp long. This improves the assignment to genomic or EST sequences. Quantification with the SAGE method and in particular with LongSAGE is problematic, e.g. it is often necessary to correct multiplication-related distortions of the quantification using biostatistical analysis methods (e.g. Emmersen, J. et al., BMC Bioinformatics 2007, 8: 92). However, these methods are based on statistical assumptions and can therefore only provide an approximate estimate of the actual relative frequency of tags in a sample.

- “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) uses the type III restriction enzyme EcoP15I as a tagging enzyme to produce 25–27 bp long tags, which offer further advantages in addition to improved assignment to other sequences.
- “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) was developed to identify transcript initiation sequences and promoters. In this method, 21 bp long fragments that are directly adjacent to the 5'-CAP sequence of an mRNA are isolated and sequenced.
- “Gene identification signature” (GIS, C. L. Wei et al., 5' Long serial analysis of gene expression (LongSA-GE) and 3' LongSAGE for transcriptome characterization and genome annotation, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004), pp. 11701–11706) Used to simultaneously determine the 5' and 3' ends of mRNAs. GIS tags are 20 bp long and can be used to identify the ends of a gene in the genome.
- “Digital karyotyping” DK; (T. L. Wang et al., Digital karyotyping, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002), pp. 16156–16161) adapts the LongSAGE protocol to obtain 21 bp tags from genomic DNA. This makes it possible to precisely determine the pathological ‘proliferation or loss of certain chromosome segments as well as insertions and foreign DNA, e.g. in cancer cells (J. T. Park et al., Notch3 gene amplification in ovarian cancer, *Cancer Res.* 66 (2006), pp. 6312-6318).
- “Methylation-specific digital karyotyping” (MSDK) (Min Hu et al., Methylation-specific digital karyotyping *Nature Protocols* 1, - 1621-1636 (2006)) applies the principle of digital karyotyping for the quantitative determination of different methylation states of two samples.
- “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) aims to identify protein-binding sequences in the genome. The principle of GIS is used to isolate and quantify tags from both ends of a DNA fragment that are bound to a specific protein (e.g., a transcription factor), which is specifically precipitated from the protein-DNA mixture together with the bound DNA by immuno-precipitation. Paired-end ditag was used to identify p53 binding sites in the human genome.
- Metagenomic analyses aim to determine the composition of a complex mixture of organisms. The same methods are used in digital karyotyping. Krause et al., Phylogenetic classification of short environmental DNA fragments, *Nucleic Acids Res.* 2008 April; 36(7): 2230–2239.

[0007] US 2007/0172973 also describes a method for the quantitative determination of nucleic acids, wherein the nucleic acids to be quantified are first sorted according to their labeling and then their quantities are determined.

[0008] In all of the methods described, the number of nucleic acids available in a sample is usually insufficient for all subsequent analyses. This is why it is almost always necessary to multiply the material before analysis.

[0009] For example, highly parallel sequencing methods (e.g., the “454” picoliter method from Roche Diagnostics, Germany, or the Solexa pyrosequencing method from Illumina, Inc., San Diego, USA) require 5–3000 ng of DNA. In addition, both methods require specific sequences (adapters) at both ends of the nucleic acid to be sequenced, which serve as primer binding sites. To ensure the presence of these specific sequences in all molecules to be sequenced, the polymerase chain reaction (PCR) is used, wherein oligonucleotides complementary to the adapter sequences are used as primers. Prior amplification is often necessary for cloning the fragments.

[0010] However, amplification, e.g. by PCR, harbors the risk that certain nucleic acids are preferentially amplified. Shorter nucleic acids are preferred, but the individual base sequence can also influence the efficiency of amplification. This means that the relative frequency of nucleic acid in a nucleic acid population can vary greatly before and after amplification.

[0011] The quantification of nucleic acid compositions is also frequently carried out without prior PCR amplification, e.g. after cloning of the nucleic acids, as is the case with the quantification of expressed sequenced tags (ESTs). Here too, the original composition of the nucleic acid population may be distorted due to differences in the fitness of individual bacteria and due to the effects of nucleic acids on the growth of the transformed cells.

[0012] The objective of the present invention was to provide a simple method for better detecting distortions in the nucleic acid compositions occurring during the amplification of nucleic acids to be quantified in a sample and for correcting these distortions during quantitative analysis.

[0013] The task is solved by labeling all nucleic acids of a nucleic acid mixture, regardless of whether they are double-stranded or single-stranded, with a "random tag" (RAT) derived from a mixture of synthetic oligonucleotides, wherein the mixture contains sufficient RATs with different sequences to ensure that the resulting random tag nucleic acid (RAT-NS) combinations are practically unique. The methods used to label nucleic acids with a RAT prior to amplification can vary greatly. After sequencing, the RAT nucleic acid combinations can be determined, resulting in a data set containing the individual sequence information of the RAT NS. RAT nucleic acid combinations that occur more than once or at least more frequently than statistically probable are therefore recognizable as copies. If all copies are eliminated from the data set, the original composition of the nucleic acid mixture prior to amplification is obtained.

[0014] Consequently, the object of the present invention is a method for the quantitative determination of nucleic acids (nucleic acid population) in a sample in which

- a) nucleic acid constructs comprising
 - at least one characteristic polynucleotide domain (TAG) of a nucleic acid present in a sample that is to be quantitatively determined, and
 - at least one artificial polynucleotide domain with a random sequence ("random tag" or RAT for short) and
 - at least one primer or polymerase binding site (PB)

wherein a characteristic polynucleotide domain (TAG) and at least one polynucleotide domain with a random sequence (RAT) form a characteristic domain combination (TAG-RAT combination) which is flanked by at least one polymerase or primer binding site (PB), are amplified in the presence of a polymerase and

- b) after amplification, at least the sequence of the polynucleotide domains with random sequence (RAT) and at least part of the characteristic polynucleotide domain (TAG) can be determined, wherein, for quantitative determination of the nucleic acid present in the sample, sequence-identical and, if necessary, additionally sequence-similar combinations of a characteristic polynucleotide domain (TAG) and a polynucleotide domain with random sequence (RAT) as copies of a characteristic polynucleotide domain (TAG) of the nucleic acid originally present in the sample can be identified.

[0015] The term "nucleic acid" as used in the present invention refers to nucleic acids present in a sample that can be determined quantitatively. The nucleic acids to be determined may be present in the sample as single strands or double strands. These are DNAs and RNAs, in particular mRNAs, microRNAs, rRNA and non-coding RNAs, as well as their cDNAs generated by reverse transcription, which can be used for gene expression analysis and analysis of non-coding RNAs. In addition, the term "nucleic acid" also includes aptamer DNA and RNA or fragments of genomic DNA, which may also be methylated. The invention further aims to quantify DNA fragments obtained by immunoprecipitation and fragments quantified by digital karyotyping and methylation-specific digital karyotyping. The invention also improves the quantification of DNA fragments for metagenomic analyses and for the analysis of fragments from subtractive DNA libraries. The term "nucleic acids" also includes single- or double-stranded polynucleotides of artificial (synthetic) origin, including those whose phosphate groups have been replaced by other atoms or whose bases have been modified and which can be amplified by polymerization and sequenced.

[0016] The term "primer or polymerase binding site (PB)" refers to a nucleic acid of known sequence which can preferably serve as a binding site for a primer, e.g. for PCR, with amplification being carried out using the corresponding primers and polymerases by enzymatic polymerization. Furthermore, known sequences can also be used for polymerase binding to initiate enzymatic amplification, such as those already contained in known plasmids or other vectors. In such a case, using an insertion sequence, the RAT-TAG combinations can be incorporated into a vector, thereby producing a nucleic acid construct within the meaning of the present invention that can be propagated by cloning. The polymerase binding site can be, for example, a promoter sequence for an RNA polymerase, such as SP6 or T7 RNA polymerase, wherein linear amplification of the nucleic acid construct by the corresponding polymerase can take place from the binding

site.

[0017] A “characteristic polynucleotide domain” refers to a so-called “TAG,” namely a nucleic acid domain whose sequence is characteristic of a nucleic acid present in the sample and to be determined quantitatively. In principle, any part of a nucleic acid can be selected as a characteristic polynucleotide domain, provided that the individual TAGs can be assigned to a specific nucleic acid present in the sample. If this assignment is not unambiguous, further methods can be used to ensure this unambiguity, e.g. with real-time PCR with specific primers. For example, conserved parts of a gene sequence can be used as a TAG, but it is preferable to use labeling polynucleotide domains with a sufficient sequence length. The characteristic polynucleotide domains should preferably have a sequence length of at least 8 nucleotides, preferably at least 12 nucleotides, and particularly preferably at least 15 nucleotides. A characteristic polynucleotide domain can also be the entire nucleic acid to be determined. However, the domain should preferably not exceed a length of 50 nucleotides, preferably 35 nucleotides, and in particular 27 nucleotides. Longer, characteristic polynucleotide domains can also be incorporated into a nucleic acid construct, although it is often sufficient to determine only part of this sequence following amplification of the constructs.

[0018] A “polynucleotide domain with a random sequence” within the meaning of the present invention is a synthetically produced (artificial) “random tag” (RAT) whose sequence allows the characteristic polynucleotide domain to be uniquely marked. RATs are usually produced commercially by “statistical” linking of the individual nucleotides. This will result in a mixture of RATs, wherein the probability of RATs with the same sequence decreases as the sequence length increases and the number of RAT molecules in the mixture decreases. The polynucleotide domains produced in this way have a random sequence within the meaning of the invention. The production of RATs with a known sequence is more complex, but this method guarantees that the sequence of the RATs produced is unique. Such RATs also constitute a “random sequence” within the meaning of the present invention. RATs can also be produced from a mixture of known, different sequences.

[0019] Single-stranded RATs can be produced as oligonucleotides or purchased. Preferably, single-stranded oligonucleotides with a random sequence (RAT) and a known sequence fraction are used to produce double-stranded RATs, wherein the known sequence fraction lies at the 3' end of the single-stranded oligonucleotide. Oligonucleotides with a complementary sequence are hybridized to the known sequence fragment as primers and converted into a double-stranded oligonucleotide containing a double-stranded RAT using a polymerase (e.g., with the so-called Klenow fragment of DNA polymerase I). Pure double-stranded RATs without additional domains can be generated by integrating a tagging enzyme site and a primer binding site, wherein the RAT is cleaved after second-strand synthesis using the tagging enzyme. Furthermore, the single-stranded RAT can preferably be provided at one end with a known sequence that can contain certain polynucleotide domains such as a tagging enzyme site, restriction site, ligation site, etc. At the other end of the RAT, a primer binding site is then preferably inserted, which can also be used for the synthesis of the double strand with the Klenow fragment. In order to provide RATs with suitable ligation sites (LS) that are complementary to the overhanging restriction interfaces, oligonucleotides can be produced that contain a primer binding site followed by a RAT and the corresponding restriction interface followed by another RAT and another primer binding site. You will obtain a construct with the following structure: PB-RAT restriction interface-RAT-PB. After the double-stranded construct has been generated, the oligonucleotides are digested with the corresponding restriction enzyme, and two RATs are obtained from each oligonucleotide, which are modified with overhanging ends for ligation.

[0020] In addition, the RATs can be produced by hybridization of two single-stranded oligonucleotides consisting of a random combination, which are hybridized with each other under stringent conditions. For RATs that are constructed from known sequences, this is the preferred method.

[0021] "Nucleic acid constructs" are understood to be polynucleotides that contain at least one characteristic polynucleotide domain that is characteristic of a nucleic acid contained in the sample and to be determined quantitatively (TAG), a polynucleotide domain with a random sequence (RAT) and at least one primer or polymerase binding site (PB). In addition, further polynucleotide domains may be contained in the nucleic acid constructs, such as linker sequences, ligation sites (LS), restriction enzyme interfaces or tagging enzyme binding sites (TES). In a preferred sequence of polynucleotide domains, the characteristic polynucleotide domain (TAG) and the polynucleotide domain with random sequence (RAT) are covalently linked to each other. The link can be established by direct ligation of the individual domains. Alternatively, the individual domains can also be linked together via linker sequences, wherein the linker sequence is formed by ligating corresponding overhanging ends. The linked TAG and RAT domains are finally supplemented by at least one primer or polymerase binding site (PB), which is preferably covalently bound either to the characteristic polynucleotide domain (TAG) or to the polynucleotide domain with a random sequence (RAT). In a preferred embodiment, the TAG domain and the RAT domain are flanked by two different primer binding sites (PB, PB'). In addition to a direct link between the primer or polymerase binding sites and the respective ends of the TAG-RAT domain combination, a link via a polynucleotide linker may also be possible. The primer or polymerase binding sites (PBs) can also be incorporated into the nucleic acid constructs by ligation. Alternatively, TAGs can also be used, especially RATs, which are already equipped with a primer binding site (PB) in advance.

[0022] In an alternative embodiment, the nucleic acid fragments prepared for enzymatic amplification may contain, in addition to at least one characteristic polynucleotide domain (TAG), two or more polynucleotide domains with random sequences (RATs). One, two, or more combinations of a TAG and a RAT can be contained in a nucleic acid construct. The orientation of the TAG-RAT combinations in the structure is irrelevant. Furthermore, individual TAG-RAT combinations can exist independently of each other, flanked by a primer or polymerase binding site (PB). However, it is also possible that two primer binding sites (PBs) include all TAG-RAT combinations in a nucleic acid construct or that a primer or polymerase binding site (PB) is located at one end of the TAG-RAT combinations.

[0023] In particular, nucleic acid constructs with the following polynucleotide domain sequence are of particular interest:

- 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; with n = 2 to 50, preferably with n = 2 to 10
 - h) PB-RAT-(TAG)_n-RAT-PB; with n = 2 to 50, preferably with n = 2 to 10
 - i) PB-RAT-(TAG)_n; with n = 2 to 50, preferably with n = 2 to 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'; with n = 3 to 50, preferably with n = 3 to 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'; with n = 0 to 10
- (PB, PB', PB'', PB''', PB'''' = primer binding sites, wherein the primer binding sites preferably have different sequences but may also have the same sequence.)

[0024] Nucleic acid constructs according to embodiments a), b), c), e), g), and h) are particularly preferred in the context of the present invention.

[0025] Nucleic acid constructs may be single-stranded or double-stranded, wherein in single-stranded nucleic acid constructs, a sequence corresponding to a primer binding site on the complementary strand is also understood to be a primer binding site, whereas in a double-stranded nucleic acid construct, both the primer binding site and its corresponding sequence on the counterstrand are contained.

[0026] The nucleic acid constructs serve as templates for their enzymatic amplification by a polymerase.

[0027] The nucleic acid constructs may also contain other oligonucleotide domains, such as linkers, ligation sites, restriction sites and tagging enzyme binding sites (TES).

[0028] In a further preferred embodiment, the nucleic acid construct or precursors of this construct contain one or more binding sites for a tagging enzyme (TE). A tagging enzyme site (TES) is a binding site for a type II or type III restriction enzyme that cuts away from its recognition site, preferably EcoP15I, MmeI, or BsmfI. By using a tagging enzyme, nucleic acid fragments of a defined length can be produced, which serve as a characteristic nucleotide domain (TAG). The tagging enzyme site (TES) must be incorporated in such a way that the required TAG is still present in the nucleic acid construct or in a construct precursor after enzymatic cleavage. The preferred binding site for a tagging enzyme (TES) is at the distal end of the PB sequence of the RAT; the integration of the TES into a nucleic acid construct or a corresponding construct precursor therefore preferably occurs in the domain sequence "PB-RAT-TES-nucleic acid." The product of the subsequent cleavage of the nucleic acid construct with the TE is therefore a polynucleotide with the domain sequence PB-RAT-TES-TAG. Alternatively, the TAG can first be generated using the tagging enzyme, and then the RAT is attached to any end of the TES-TAG domain sequence. If, for example, EcoP15I is used as a tagging enzyme, the presence of two head-to-head oriented recognition sites (tagging enzyme sites) with the sequence CAGCAG is required, which include the interface of the tagging enzyme and are preferably located at a distance of less than 3000 bp. The preferred domain sequence is therefore "PB-RAT-TES nucleic acid TES" when using a type III restriction enzyme such as EcoP15I. The product of the subsequent cleavage of the nucleic acid construct with EcoP15I as TE is therefore a polynucleotide with the domain sequence PB-RAT-TES-TAG, and a polynucleotide with the domain sequence TES-Tag, of which the PB-RAT-TES-TAG construct is preferred for characterizing the nucleic acid. A major advantage of using EcoP15I is that particularly long TAGs with 25–27 base pairs are obtained.

[0029] In another preferred form of application, both ends of the nucleic acid molecule to be characterized are provided with identical or different PB-RAT-TES constructs. The cleavage of the construct with TE then results in two molecules with the domain sequence PB-RAT-TES-TAG, wherein the PB and TES domains may have the same or different sequences. In this case, two PB-RAT-TES-TAG constructs can be used to characterize the nucleic acid molecule.

[0030] After cutting with the corresponding tagging enzyme, further polynucleotide domains, such as RATs and PBs, can then be attached to the resulting TAG, so that the following particularly preferred domain sequences are obtained in the resulting nucleic acid constructs:

- 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;
- with m = 0 or 1 independently of each other for each m, where at least one m = 1
- 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';

with n = 3 to 50, preferably with n = 3 to 10 and m = 0 or 1 independently of each other for each m,

n') PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''; with m = 0 or 1 for each m independently of each other, wherein at least one m = 1,

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''''';

with m = 0 or 1 for each m independently of one another, wherein at least one m = 1,

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'

with n = 0 to 10 and with m = 0 or 1 for each m independently of one another, wherein at least one m = 1, (PB, PB', PB'', PB''', PB''', PB'''' = primer binding sites, wherein the primer binding sites preferably have different sequences but may also have the same sequence).

[0031] The nucleic acid constructs can be produced using circular nucleic acid construct precursors, wherein preferably at least one end of a nucleic acid is modified with a marker and, if desired, the other end is also modified with a marker or adapter, and circularization then takes place. Preferably, at least one of the existing markers or adapters also has a tagging enzyme binding site (TES). Two TES are particularly preferred, especially for the tagging enzyme EcoP15I, which cuts off a TAG on both sides of the circular construct, thereby creating a linear nucleic acid construct within the meaning of the present invention that contains two TAGs at both ends.

[0032] Further preferred nucleic acid constructs, which can be generated, for example, from circular construct precursors and which are of particular interest for the simultaneous quantification of both ends of nucleic acid molecules because they allow the origin of the TAGs from the same molecule to be recognized, are

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' or

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''' = primer binding sites, wherein the primer binding sites preferably have different sequences but may also have the same sequence.)

[0033] All nucleic acid constructs shown in the present description of the invention may have further sequence segments, such as linkers, restriction sites, ligation sites, etc., in addition to the domain sequences according to the invention, which, however, are generally not explicitly mentioned in the domain sequence of the constructs.

[0034] Randomly amplified polynucleotide (RAT) domains are suitable for marking characteristic polynucleotide domains (TAGs). A "marker" is therefore understood to mean polynucleotides that contain a primer binding site and a RAT, but no TAG. In addition, markers can contain other oligonucleotide domains such as tagging enzyme binding sites (TES) and linker and ligation sites (LS). A ligation site is the end of the marker intended for ligation, which is located distally to the PB.

[0035] For the production of the nucleic acid constructs intended for amplification, it is also useful to combine certain oligonucleotide domains in an "adapter" in addition to the markers. Adapters within the meaning of the present invention are always provided with a primer binding site and may contain further oligonucleotide domains, in particular tagging enzyme binding sites, linkers or ligation sites, but no RATs or TAGs.

[0036] The markers and adapters preferably contain different primer binding sites, which should preferably be used in the highly parallel sequencing methods according to the manufacturer's instructions. These are located at the ends of the markers or adapters distal to the TAG.

[0037] The individual oligonucleotide domains, such as PB, TAG and RAT, or the markers or adapters can also be provided with further known modifications using the methods customary in the prior art, which for example

- enable binding to a solid phase, for example by biotinylation of the oligonucleotides and binding to a solid phase provided with streptavidin or contain "capture sequences". Sequences can be used as "capture nucleotide sequences" which, for example, are bound to a solid phase and can hybridize with their corresponding counterstrands which are in solution,
- prevent ligation on one side of an oligonucleotide, e.g. amino-C7 modifications or "reverse nucleotides (nucleotides with 3'-5' orientation instead of 5'-3'-orientation)", wherein the end to be protected from ligation is the end of the marker or adapter distal to the TAG,
- enable or facilitate ligation on one side of an oligonucleotide, e.g. by phosphorylation of the 5' end of the corresponding oligonucleotide, or e.g. by the generation of overhanging restriction interfaces ("sticky ends"),
- facilitate hybridization or make dehybridization of the double strand more difficult, e.g. by the use of "locked nucleotide acids"
- include restriction interfaces

[0038] These modifications can be integrated into the single-stranded oligonucleotides. Adapters can be hybridized with complementary oligonucleotides to form double-stranded oligonucleotides. Markers can also be produced, if necessary, but it is preferable to produce the marker double-stranded using a polymerase, in particular the Klenow fragment of DNA polymerase I, as described above.

[0039] With the aid of markers, nucleic acid fragments with a sequence that is characteristic of a nucleic acid present in the sample (TAG) can be uniquely identified, i.e., each RAT-TAG combination is highly likely to be unique. In order to achieve such unique identification, the random sequence nucleic acid domain (RAT) should preferably be long enough to provide an excess of markers with different RATs in relation to the number of nucleic acids present in the sample. Preference is given to markers containing RATs with more than twice the amount relative to the nucleic acid fragments or nucleic acids to be labeled. In individual cases, however, a marker number corresponding to the number of nucleic acids to be determined may also be sufficient, especially if all RATs are guaranteed to differ in their sequence. Preferred sequence lengths for the RATs contained in the nucleic acid fragments are between 2 and 100 nucleotides, preferably between 4 and 50 nucleotides, with particularly preferred RATs having a total length between 4 and 15 nucleotides. The entire random sequence may be contained in a marker domain or may be divided among several marker domains or among domains of different markers in a nucleic acid construct. With the described total lengths of the nucleic acid domains with random sequences (RATs), it is theoretically possible to generate a large number of markers with different RATs, wherein the marker ensemble used usually contains only a small portion of the conceivable RATs resulting from the sequence length, so that the probability of randomly identical RATs occurring in two markers is very low. Consequently, the sequence length of the RATs can be individually adapted to the number of distinctive nucleic acid fragments to be labeled. For example, for a gene expression analysis in which the mRNAs contained in a tissue sample are to be quantitatively determined, marker mixtures with more than 102 markers, particularly preferably between 10^4 and 10^8 markers, are preferably used.

[0040] Even if individual markers should actually contain identical RATs, these only distort the quantitative determination if they are linked to a sequence-identical or sequence-similar TAG. Since all individual TAG-RAT combinations for the individual nucleic acids present in the sample and to be quantified in a mixture prepared for amplification should only be present once, sequence-identical combinations of a characteristic polynucleotide domain (TAG) and a polynucleotide domain with a random sequence (RAT) can therefore be identified as copies of a characteristic polynucleotide domain of a nucleic acid originally present in the sample.

[0041] Since polymerization errors can also occur during the amplification of nucleic acid constructs and the determination of the nucleic acid sequence is not free of errors, it is advantageous to also eliminate sequence-similar RAT-TAG combinations during evaluation. What counts as "sequence-like" combinations can basically be decided freely.

[0042] The frequency with which such errors occur depends on the length of the TAGs and RATs, the amplification conditions, and the sequencing method. Sequence-like copies are preferably considered to be RAT-TAG sequences that have a maximum of one (1) deviation per 10 nucleotides. Sequence-like copies are preferably defined as RAT-TAG combinations that have no more than one deviation per 15 or 20

nucleotides. For example, sequence-like RAT-TAG combinations with a sequence length of 40 base pairs should have a maximum of 4 sequence deviations if error-tolerant conditions are used. Provided that amplification errors and sequencing errors can be largely eliminated, it is advantageous to lower the error threshold for “sequence-like” RAT-TAG combinations.

[0043] Fig. 1 shows an example of four RAT-TAG combinations with the same TAG. Combinations 1, 2, and 3 are different; combination 4 is a copy of 3. The TAG therefore appears three times in the sample. For quantification, only one combination would be counted, e.g. 3, while combination 4 is eliminated from the data set.

[0044] The nucleic acid containing the TAG is preferably double-stranded, e.g. as cDNA or DNA, and is ligated downstream of the double-stranded construct(s) (marker, adapter). For linear amplification of RNA, however, the nucleic acid to be amplified can also be present as a single strand. Single-stranded RNA can be labeled using RNA ligase, e.g. T4 RNA ligase, prior to amplification.

[0045] The following describes some strategies for producing nucleic acid constructs from double-stranded DNA. Due to the modular nature of the individual method steps for producing the constructs, nucleic acid constructs with other domain sequences can also be generated in a similar manner, or identical nucleic acid constructs with an alternative sequence of method steps can be obtained.

[0046] The individual method steps which are advantageous for producing the nucleic acid constructs are

- the binding of single- or double-stranded oligonucleotides, in particular of markers and adapters, or of the nucleic acids to a solid phase, and their cleavage
- the ligation of the individual markers, adapters, and nucleic acids, and
- the digestion of nucleic acid strands with a restriction and/or tagging enzyme

[0047] In a preferred embodiment for generating quantitatively determinable, double-stranded nucleic acid constructs containing, for example, TAGs from cDNA or genomic DNA, double-stranded nucleic acids are provided with a marker or adapter which can be bound to a solid phase.

[0048] Binding to a solid phase has the advantage that, for example, nucleic acids that are not ligated or not required can be removed after restriction digestion or required nucleic acids can be removed from the solid phase. However, binding nucleic acids to a solid phase is not necessary for some methods.

[0049] Preferably, markers or adapters for this purpose are biotinylated and can be bound to a streptavidin-coupled solid phase. The nucleic acids to be analyzed in a mixture can be ligated with such markers or adapters that are biotinylated at the opposite end of the end intended for ligation (ligation site = LS). This can be done, for example, after digestion of the DNA to be quantified, with a restriction enzyme at a restriction interface or also at the smooth ends of the nucleic acid. Biotinylated markers or adapters can also be used in cDNA, preferably by using biotinylated single-stranded oligonucleotides that have a poly-T sequence at the 3' end and are integrated by first- and second-strand synthesis starting at the 3' end of the cDNA (the 3' end of the cDNA is understood to be the end corresponding to the 3' end of the mRNA. It is characterized by an oligo-A/T domain).

[0050] In a preferred embodiment, the nucleic acid (cDNA, DNA) is digested with one or more frequently cutting restriction enzymes. Frequently cutting restriction enzymes are, for example, NlaIII, Hsp92II, FatI, BfaI, MaeI, XspI, HpyCH4IV, MaeII, Tail, TscI, AluI, TaqI, BfuCI, Bsp143I, BstENII, DpnII, Kzo9I, MboI, NdeI, Sau3AI, BstKTI or Csp6I. These enzymes produce nucleic acid fragments with an average length of 200 to 300 base pairs. The nucleic acids of the mixture to be quantified can be bound to a solid phase before or after this digestion, e.g. after ligation with biotinylated markers or adapters. As an alternative to digestion, the nucleic acids to be determined in the mixture can be broken down by shearing before binding to the solid phase. Fragments that are not bound to the solid phase can be removed by washing.

[0051] The bound fragments can preferably be relabeled by ligation with a marker or adapter, which should not be biotinylated. If the nucleic acids to be quantified have been previously digested with a restriction enzyme, markers or adapters can be provided with ends corresponding to the restriction interface and ligated with the digested nucleic acids to be quantified. The other end of the adapter or marker can be protected against ligation, e.g. by installing an Amino C7 modification in the end of the marker or adapter distal to the LS. After mechanical shearing, the nucleic acids to be quantified can also be provided with non-biotinylated, suitable markers or adapters after filling or removing protruding single strand ends (“sticky ends”), e.g. with the Klenow fragment.

[0052] The nucleic acids of the mixture to be quantified are preferably in the form: PB-TAG-RAT-PB', PB-RAT-TAG-RAT or PB-RAT-TAG-RAT-PB' and can be amplified e.g. by PCR and sequenced e.g. by highly parallel sequencing methods.

[0053] For quantitative analysis of gene expression, it is generally sufficient to use the cDNA ends (Torres T. T. et al., Gene expression profiling by massively parallel sequencing, Genome Res. 2008 18: 172–177) as TAGs. For this purpose, mRNAs are converted into a cDNA using, for example, a biotinylated poly-T primer (PB1 = adapter) and bound to a solid phase. This is followed by digestion with a restriction enzyme (anchoring enzyme), which often has a cutting action. This is followed by ligation with a PB'-RAT marker containing a second primer binding site (PB2), wherein the end distal to the ligation site is protected from ligation. The resulting nucleic acid construct is therefore PB2-RAT-TAG-PB1. The nucleic acid constructs obtained are amplified by PCR and then sequenced.

[0054] Another possibility for obtaining nucleic acid constructs that are particularly suitable for highly parallel sequencing methods is the use of biotinylated markers or adapters, whose binding to a solid streptavidin phase and the extraction of single-stranded constructs after denaturation. The markers or adapters are 5'-phosphorylated distal to the primer binding site and protected from ligation at the end where the primer binding site is located, e.g. with an amino C7 modification. The constructs c, d and e, which are particularly suitable for pyrosequencing, are obtained by ligating the TAG, which can be, for example, a cDNA or a genomic DNA fragment, with markers and adapters, e.g. via specific restriction interfaces or blunt ends, wherein one of the individual strands of the marker, e.g. PB-RAT-LS or adapter, e.g. PB'-LS, which is composed of two individual strands, is biotinylated. After ligation, the unwanted by-products PB'-TAG-PB and PB-RAT-TAG-RAT-PB are also produced. These are now either biotinylated on both strands or not biotinylated at all. After binding to a streptavidin phase, the unbound fragments are first removed by washing. If the double-stranded constructs are denatured, the fragments with two biotinylated adapters remain on the solid phase. Only those nucleic acid constructs with the desired formation dissolve as single strands and can thus be recovered. The single-stranded constructs can be regenerated into double strands with the help of a polymerase. This method is versatile and can in principle be used for all double-stranded DNA nucleotides.

Extraction of TAGs using a tagging enzyme:

1) Obtaining a tag labeled with a RAT from a cDNA:

[0055] In a further embodiment, the nucleic acids of a mixture to be quantitatively determined are provided with nucleic acid constructs (markers and adapters) as described above, which contain a tagging enzyme binding site (TES) and are bound at one end to a biotin molecule (e.g., TES-RAT-PB-biotin). This allows the nucleic acids to be bound to a solid phase, e.g. streptavidin-coupled magnetic beads. After digestion with one or more frequently cutting restriction enzymes or scissors, the construct can be cut with a tagging enzyme, such as BsmFI or Mmel, after it has been bound to a solid phase.

[0056] It should be noted here that fragments cleaved by a tagging enzyme have the same length, provided that the markers and adapters used are also of the same length and that all molecules have the same sequence of domains, as is preferably the case according to the present invention. The nucleic acid fragments cleaved by the tagging enzyme can thus be easily separated and isolated by size, e.g. electrophoretically or by HPLC. Even after ligation with a marker or adapter of a certain size, the ligation products can be isolated based on their size. The corresponding specific amplification products can also be isolated in this way.

[0057] If a Type III enzyme such as EcoP15I is used as a tagging enzyme, an adapter or marker containing a second TES must first be ligated to the free end of the nucleic acid strands fixed to the solid phase. After digestion with the tagging enzyme and removal of the cleaved products that are not bound to the solid phase, any protruding ends can be filled with the Klenow fragment, for example, or cleaved with a single-strand-specific exonuclease so that the tagging enzyme interface has blunt ends. Markers or adapters can then be bound to the fixed nucleic acid strand. Alternatively, adapters or markers with “sticky ends” can be used, which are complementary to the end created by the tagging enzyme.

[0058] Depending on the positioning of the marker or adapter, the products PB-TAG-TES-RAT-PB' or PB-RAT-TAG-TES-RAT-PB' are preferred.

[0059] Nucleic acid constructs can be produced both using a solid phase and in solution. In a preferred embodiment, the nucleic acids or nucleic acid fragments contained in a sample are ligated at both ends with markers or adapters. The ends can be defined in advance using a restriction enzyme and provided with suitable markers or adapters. The resulting construct has the following domain sequence:

PB nucleic acid PB

or:

PB-RAT nucleic acid RAT-PB

[0060] If this construct is digested with an enzyme that frequently cuts, further interfaces are often created in the nucleic acid. These can be used for ligating another adapter or marker. The following products are created:

PB-RAT-TAG-PB

or

PB-TAG-RAT-PB

[0061] If adapters or markers containing tagging enzyme recognition sites are used in the first step, the following is produced:

PB-RAT-TES-Nucleic acid-TES-RAT-PB'

or:

PB-TES nucleic acid TES PB'

[0062] After digestion with the tagging enzyme, two nucleic acid fragments containing a TAG can be cleaved from the nucleic acid strands obtained. A marker or adapter is preferably attached to the interfaces of the tagging enzymes. In addition to others, the following nucleic acid constructs will be produced, which can be isolated by means of gel electrophoresis or HPLC, for example:

PB-RAT-TES-TAG-PB'

and

PB'-RAT-TES-TAG-PB

or:

PB-TES-TAG-RAT-PB'

and

PB'-TES-TAG-RAT-PB

[0063] The formation of two nucleic acid constructs for each nucleic acid contained in the sample enables internal control during the quantitative determination of the nucleic acids.

[0064] It can be seen that the nucleic acid constructs required for amplification, which contain the characteristic TAG-RAT combinations, can be produced in a variety of ways. The nucleic acid constructs can be assembled in a modular manner, e.g. the individual method steps for binding or generating individual polynucleotides with the different functional domains (TAG, RAT, PB, TES) can be easily combined to obtain the desired nucleic acid constructs. This makes the method according to the invention versatile and not limited to a specific processing of the nucleic acids contained in the sample.

[0065] The individual, modular method steps are explained in more detail below using examples.

1) Preparation of an RNA-containing sample for the production of cDNA tags:

[0066] Total RNA is extracted from a biological sample using a commercially available kit, e.g. with Trizol reagent (Invitrogen Corp., Carlsbad, USA). The mRNA is isolated from the total RNA using a commercially available kit, e.g. "Oligtex Midi Kit" (Qiagen N. V., Venlo, Netherlands). The cDNA is produced by reverse transcription using a 5'-biotinylated poly-T-oligonucleotide, e.g. with the "cDNA synthesis system" (Invitrogen Corp.). The product is converted into double-stranded DNA ("cDNA synthesis system," Invitrogen Corp.) and dissolved in a suitable buffer. In a particularly preferred variant of the method, the poly-T-oligonucleotide used should contain the recognition sequence for EcoP15I CA- GCAG or CTGCTG, e.g. 5'-biotin-PB-CAGCAGTTTTTTTTTTTTTTTTTTT-3' (SEQ ID No. 1).

2. Enzymatic digestion of nucleic acids present in the sample:

[0067] Double-stranded DNA is digested in a reaction solution with a frequently cutting restriction enzyme ("anchoring enzyme"), e.g. NlaIII, in a buffered solution. After digestion, the DNA obtained is extracted, precipitated with alcohol, and resuspended in a suitable buffer. The DNA solution is then mixed with streptavidin-coated particles. The particles are incubated at room temperature in a suitable buffer to allow the biotinylated cDNA to bind to the streptavidin-coated particles. The DNA bound to the particles is washed and absorbed into a suitable buffer.

3) Production and ligation of polynucleotide domains, adapters, and markers:

[0068] Adapters are composed of two complementary single-stranded oligonucleotides. The two single-stranded oligonucleotides are dissolved in a suitable buffer and hybridized with each other to form a double strand. The marker containing the RAT is also initially synthesized as a single strand and a double strand is synthesized from this template as described above using the Klenow fragment. Adapter and marker are protected from unwanted ligation on the side distal to the LS by the incorporation of, for example, an amino C7 modification of the 3' side of the corresponding oligo, while ligation is enabled at the ligation site by 5'-phosphorylation. The double-stranded marker or adapter can now be covalently linked with a ligase, e.g. T4 ligase (Invitrogen Corp.) in a suitable buffer to a 3'-end of another polynucleotide or a nucleic acid from the sample or a fragment thereof, which contains the TAG and can be bound to a solid phase. If markers and adapters are to be ligated, they should preferably be added in a sufficient quantity so that all free ends of the nucleic acid strands to be constructed can also be ligated with the polynucleotides.

4) Cutting with the tagging enzyme:

[0069] In a suitable buffer, the nucleotides to be quantified, which are provided with an adapter or marker containing a TES, are incubated with a tagging enzyme, e.g. BsmFI, Mmel, EcoP15I. The separated nucleic acid fragments are isolated either after electrophoresis or HPLC or are bound to a solid phase (when using appropriate markers or adapters). If additional markers or adapters are to be attached to the tagging enzyme interface and the interface is not smooth, the overhanging end can be filled in or cleaved off with an exonuclease. The Klenow fragment of DNA polymerase I, for example, can be used for this purpose.

5) Amplification:

[0070] PCR-based amplification methods are preferred for the replication of nucleic acid constructs. Alternatively, unidirectional amplification or cloning methods can be used to replicate the nucleic acid constructs.

6) Sequencing:

[0071] The sequence of individual nucleic acid fragments obtained by amplification in a mixture is preferably determined by suitable sequencing methods. Such methods are offered commercially, e.g. "454" picoliter method, with the GSFLX system (Roche Diagnostics, Germany), Solexa method, (Illumina Inc., San Diego USA); SOLid method (Applied Biosystems Inc, Foster City; USA); HeliScope™ Single Molecule Sequencer, (Helicos Corp., Cambridge, USA). Using these methods, it is now possible to sequence up to several million sequences in parallel. A mixture of amplified nucleic acid constructs can be sequenced in parallel without further separation or without an intermediate cloning step. However, it is also possible to assemble the nucleic acid fragments obtained into concatemers, which can then be cloned, processed, and sequenced using conventional methods (e.g., according to Sanger).

[0072] Since the quantitative evaluation of the nucleic acid constructs produced is carried out by sequencing, the part of the constructs to be sequenced should not exceed a length of 400 nucleotides in a preferred embodiment. Preferred nucleic acid constructs therefore have a length between 15 and 150 base pairs, particularly preferably between 25 and 75 base pairs. However, longer sequences up to the entire nucleic acid sequences can also be sequenced.

[0073] Brief description of the figures illustrating examples of individual possible embodiments of the invention:

[0074] Fig. 1 shows an example of four RAT-TAG combinations with the same TAG.

[0075] Fig. 2 schematically shows the production of nucleic acid constructs from mRNA for quantitative analysis of gene expression.

[0076] In Fig. 2, steps 1.1 through 1.3 show linear amplification using a T7 promoter (PB'). For this purpose, a marker with the domain sequence PB'-PB-RAT-Poly-T, where PB' contains the promoter sequence for a T7 or SP6 RNA polymerase, is hybridized with the poly-A end of the mRNAs in the sample. The free 3' end of the poly-T sequence then serves as the starting point for reverse transcription (first strand synthesis: Fig. 2, step 1.1). A double-stranded cDNA is produced from this by second-strand synthesis, e.g. using oligohexamers (step 1.2). Linear amplification can then be performed on the cDNA using T7 polymerase to produce antisense RNA (aRNA) (step 1.3). The aRNA obtained can then be transcribed into the corresponding DNA first strand using "random hexamers" as primers by reverse transcription (step 1.4), yielding the corresponding amplified, single-stranded cDNA. The corresponding DNA strand can then be amplified using the primer binding site (PB) (step 1.6.1), wherein biotinylated primer binding sites can also be used (step 1.6.2). The resulting double-stranded constructs can then be further analyzed, e.g. as shown in Fig. 3 (step 3-5). Fig. 2, steps 2-4, shows mRNA amplification using the SMART method.

[0077] Fig. 3 schematically shows the production of nucleic acid constructs from genomic DNA using digital karyotyping as an example. To this end, the genomic DNA present in a sample is digested using a methylation-insensitive restriction enzyme (Fig. 3, step 1). The fragments obtained preferably have overhanging ends so that biotin-PB-RAT markers can be ligated to the corresponding restriction interfaces of the genomic DNA, which have suitable ligation sites distal to the primer binding site (step 2). The nucleic acid construct precursors obtained can then be digested with another restriction enzyme (step 3). After binding the marker-containing fragments to a streptavidin matrix (e.g., streptavidin-coated magnetic particles), an adapter containing a second primer binding site (PB') can then be ligated to the bound fragments via the corresponding overhanging ends ("sticky ends"). The nucleic acid constructs obtained in this way can then be amplified and sequenced. In methylation-specific DK, the first enzyme is replaced by a methylation-sensitive enzyme. In an alternative embodiment shown in Fig. 4, steps 5-8, markers containing tagging enzyme binding sites, which in this case are biotinylated, and adapters for constructing the nucleic acid constructs can also be used. The nucleic acid constructs shown in step 5 will then be contained. Short TAGs are then cut from the nucleic acid by adding the tagging enzyme (step 6). If these have protruding ends, the fragments obtained are then filled with the Klenow fragment, for example. Now, the adapter-containing fragments in solution can be ligated to the corresponding nucleic acid construct via their blunt ends (step 7.1) using markers and then amplified. The amplified nucleic acid construct can then be purified by electrophoresis and sequenced, for example. The fragment bound to the solid phase can also be further processed to form the desired nucleic acid construct by ligating the smooth end with an adapter containing a second primer binding site (step 7.2). The bound nucleic acid construct can be amplified directly and then sequenced. However, the fragments in solution containing the adapter and a TAG and the bound fragments containing the marker and a TAG can also be ligated together to form ditags (step 6.2) before amplification

and sequencing of the obtained nucleic acid construct.

[0078] The applications of the invention described below demonstrate the wide range of possible uses for the invention. The following will be described in detail:

- 1) RAT labeling of nucleic acids prior to their amplification by linear amplification:
- 2) Constructs for the quantification of cDNA,
 - 2.1.) Constructs for quantifying 3' ends of cDNA,
 - 2.2.) Constructs for quantifying 5'-ends of cDNA,
 - 2.3.) Constructs for the simultaneous quantification of 5' and 3' ends of cDNAs,
 - 2.4.) Constructs for the simultaneous quantification of 3'- and 5'-ends of the same cDNA using a circularization method,
- 3) Constructs for quantifying cDNA that was not produced from poly-A-containing RNA
- 4) Constructs for quantifying genomic DNA fragments,
 - 4.1.) Constructs for quantifying genomic DNA previously reacted with bisulfite,
 - 4.2.) Constructs for quantifying both ends of genomic DNA fragments using a circularization method.

- 1) RAT labeling of nucleic acids prior to their amplification by linear amplification:

[0079] The aim of "linear amplification" is to amplify nucleic acids, e.g. to obtain sufficient material for a gene expression study, in such a way that the proportional composition of a nucleic acid sample is preserved. There are various methods for doing this. For example, antisense RNA (aRNA) is first produced from an mRNA to which a single-stranded marker containing a RAT has been bound, using T7 or SP6 RNA polymerase (US 6,916,633) or using an RNA primer containing a T-7 promoter Dafforn et al, Biotechniques. 2004 Nov; 37(5): 854-7 linear. However, sequence-related deviations in copy number also occur during "linear amplification" of RNA (e.g., Caretti et al. 2008, J. Cellular Biochemistry 103: 556–563), which can be detected by incorporating a single-stranded RAT. In an alternative method, for example, a single-stranded marker containing a RAT can be attached to the 5' end of mRNAs contained in the sample, followed by linear amplification, e.g. from the 3' end, with poly-T primers using a reverse transcriptase.

2) Constructs for quantifying cDNA

[0080] Linear amplification is not a necessary prerequisite for the quantification of cDNAs produced from mRNA with a 3'-poly-A end. However, it is useful when mRNA is only present in small quantities. Therefore, constructs that enable linear amplification of cDNA via an RNA step are described first. The following sections describe methods that start with nucleic acids available in sufficient quantities, whether or not these have been produced by amplification. This is followed by a description of methods that enable the quantification of tagging nucleic acid domains (TAGs) from nucleic acids using RATs.

2.1. Constructs for quantifying 3'-ends of cDNA

[0081] For amplification, the mRNA is first converted into cDNA using an oligo-dT-containing primer. For subsequent amplification with a polymerase, such as SP6, T3 or T7 RNA polymerase, this primer also contains the promoter sequence for the polymerase. The subsequent double-stranded cDNA then has the domain structure "cDNA-oligo-A/T-promoter." Changes in the nucleic acid composition can be detected if a RAT is integrated between the promoter and the DNA to be amplified. The resulting domain structure in this case is "cDNA-Oligo-A/T-RAT-Promoter,"

wherein the T7 promoter sequence can also serve as a primer binding site. However, a separate primer binding site (PB) can also be incorporated next to the promoter (polymerase binding site), resulting in the following domain structure:

"cDNA oligo A/T RAT PB promoter,"

[0082] For quantification without prior direct amplification, e.g. with the aid of a polymerase such as SP6, T3 or T7 polymerase, a construct is selected that contains a primer binding site (PB) and/or a TES instead of the respective promoter. The corresponding domain structure is then: cDNA-Oligo-A/T-TES-RAT-PB

[0083] The following two preferred methods can also be used to quantify TAGs obtained from the 3'-ends of cDNA containing the poly-A portion.

[0084] For the quantification of cDNA ends in a preferred method, mRNA is transcribed into cDNA using a poly-T-containing marker, poly-T-RAT-PB-biotin, by means of first- and second-strand synthesis. The cDNA is then digested with one or more frequently cutting enzymes (anchoring enzymes). The marker-containing fragments are then bound to a streptavidin-modified matrix such as magnetic particles. The unbound fragments are removed by washing. This is followed by ligation with an adapter containing a ligation site at the end distal to the primer binding site, which is protected from ligation by an amino-C7 modification, wherein the overhanging ends are used for ligation. After removing the unligated adapters, the obtained nucleic acid constructs can be amplified and sequenced. Alternatively, a marker can be used instead of the adapter in the procedure described above. In this case, it is not necessary to transcribe the mRNA with a poly-T oligonucleotide containing RAT.

[0085] One of the preferred methods for quantitative analysis of gene expression is known as the "Super-TAG" method. For this purpose, mRNA is transcribed into cDNA using a 3'-biotinylated poly-T oligonucleotide with an EcoP15I binding site (EcoP15ITES) upstream of the poly-T sequence. This is followed by digestion with a frequently cutting restriction enzyme (anchoring enzyme), e.g. NlaIII. The cut cDNA is bound to streptavidin-coated magnetic particles via the biotin group. Alternatively, the oligo-dTs can be bound to the particles prior to cDNA synthesis, or digestion with the anchoring enzyme can take place only after binding to the particles. The unbound fragments are removed by washing. The next step is ligation with a double-stranded PB'-EcoP15 ITES oligonucleotide (adapter). The nucleic acid strands obtained are then incubated with EcoP15I and the cleaved fragments are isolated after gel electrophoresis. The adapter TAG constructs can now be ligated with a marker. The nucleic acid constructs obtained have the following domain sequence: PB'-EcoP15ITES-TAG-RAT-PB.

[0086] This is followed by PCR amplification and sequencing of the amplified constructs.

2.2. Constructs for quantifying 5'-ends of cDNA

[0087] For accurate quantification of the 5' ends of cDNAs, RATs can be attached to the 5' CAP sequence of the mRNA. The 5' CAP sequence consists of a guanine nucleotide that is bound to the mRNA by an unusual 5' to 5' triphosphate bond. To bind a RAT, for example, a procedure analogous to the production of full-length cDNA can be used (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). RNAs are treated with "alkaline phosphatase", for example. Transcripts with intact 5'-CAP sequences are not dephosphorylated by these enzymes, while broken mRNA molecules and non-coding RNA, which do not have such a CAP sequence, are dephosphorylated and end with a 5'-OH group instead of the phosphate group, to which RNA polymerase cannot bind RNA oligonucleotides. In the second step, the CAP sequence is enzymatically removed, releasing the 5'-phosphate group, which is then available for ligation of an RNA oligonucleotide. This oligonucleotide of known sequence later serves as a primer binding site (PB) for the second strand synthesis of the cDNA. To insert a RAT into this construct, the sequence of the RNA oligonucleotide should be chosen so that the subsequent RNA construct has the domain sequence 5'-PB-RAT-mRNA-3'. Optionally, a (T7 polymerase) promoter sequence can be inserted into this construct, which can then serve as a polymerase binding site and simultaneously as a primer binding site (PB). The corresponding domain structure is then
'5'-(T7)promoter-RAT-mRNA-3'

[0088] Another method for adding a primer to the 5' end of an RNA is to use the "Switch Mechanism At the 5'-end of Reverse Transcript" (SMART, TAKARA, Seta Otsu, Japan). During the first strand synthesis of the cDNA, the reverse transcriptase forms an overhang consisting of several (usually three) cytosine bases on the 3' side of the newly formed DNA strand (see also **Fig. 2**, steps 2 through 4). This is available as a binding site for an oligonucleotide primer. Here, for example, to insert the RAT, an oligonucleotide with a primer binding site at the 5' end, followed by a RAT which in turn ends with three consecutive guanine bases at the 3' end, can be used as a primer for the second-strand synthesis of the cDNA.

[0089] For many purposes, it is sufficient to obtain only one TAG from the 5'-end of a cDNA, which can then be used to quantify the original mRNA. For this purpose, as described above, a nucleic acid construct containing at least one PB (or a promoter sequence, such as from SP6, T3 or T7), a RAT and a TES is bound to the 5'-phosphate group released after dephosphorylation and removal of the CAP site. The corresponding single-stranded RNA construct then preferably has the form 5'-biotin-PB/promoter-RAT-TES-mRNA.

[0090] At the same time, a DNA construct containing a poly-T sequence can be bound to the poly-A sequence at the 3'-end of the mRNA by "hydrogen bonding" between the poly-A and oligo-T sequences of the construct. This construct contains at least one oligo-T domain, but may also consist of additional domains (see below). Starting from this double-stranded end, reverse transcription takes place, followed by second-strand synthesis. The resulting double-stranded product then has the domain structure

5'-biotin-PB/promoter-RAT-TES-cDNA-oligo-A/T-3'

and can be bound to a solid phase via the biotin group. If the TES is the recognition sequence of a type II restriction enzyme, the product can be cut immediately. While the unbound part of the cDNA can be washed away, a construct with the domain sequence

5'-biotin-PB/promoter-RAT-TES-TAG.-3'

remains bound to the solid phase, wherein the TAG, which in this case has a defined length, characterizes the cDNA.

[0091] Alternatively, a TAG can also be obtained in accordance with the invention by mechanical fragmentation of the cDNA or treatment with a non-specific endonuclease or exonuclease. However, it does not then have a defined length by definition. In this case, there is no need to introduce a TES domain into the construct. The domain sequence

5'-biotin-PB/promoter-RAT-TAG

is sufficient to characterize the cDNA.

[0092] A further primer binding site (PB), an adapter or a marker can be bound to the 3'-ends released after digestion or fragmentation. This construct, which has a defined length when a tagging enzyme is used, can then be amplified starting from the two PBs it contains.

[0093] If the tagging enzyme is a type III endonuclease such as EcoP15I, a construct containing a PB plus a TES is preferably added to the interface of the restriction enzyme or to the free end resulting from the unspecific endonuclease or mechanical disruption, which is distal to the biotin group, in such a way that a construct of the form

5'-biotin-PB/promoter-TAG-TES-PB-3'

is formed. In addition, a RAT can also be inserted here. The construct then has the following form:

5'-biotin-PB/promoter-TAG-TES-RAT-PB-3'

[0094] Such constructs can be amplified by PCR, identified on the basis of their defined size, and isolated (gel electrophoresis, HPLC).

[0095] In a particularly preferred embodiment of the manufacturing method, streptavidin-coated magnetic particles are preferably used as the solid phase. Streptavidin-coated reaction vessels are also a preferred alternative. The particle-bound end of the nucleic acid strands produced must be protected against a ligation reaction.

2.3.) Constructs for the simultaneous quantification of 5'- and 3'-ends of cDNAs

[0096] Constructs of the form described above:

5'-biotin-PB/T7 promoter-RAT-TES-cDNA-oligo-A/T-3' or other previously described constructs allow, in principle, only one TAG to be obtained from one of the cDNA ends. In special cases, however, it is desirable to obtain two TAGs from the 5' and 3' ends of the same cDNA, which can then be recognized as such. For this purpose, a circular product of the form

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

is preferably prepared, wherein PB and PB' preferably do not have the same sequence. After PB and PB' have been bound to the ends of the cDNA, the cDNA ends with known sequences can be converted into a circular molecule by ligating their PB ends under suitable conditions. These circular molecules can now be amplified by PCR starting from the PB domains. In order to determine the original composition of the cDNA population using RATs, at least one end can be labeled with a marker containing a PB domain. This can be done via the CAP site and/or the poly-A end, as described above, wherein a construct of the form

5'-PB-RAT-cDNA-3'

is formed before circularization to the 5'-end of the cDNA after removal of the CAP site as described above.

At the 3'-end, a construct of the form

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

or

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

is formed. PB and PB' can have complementary, overhanging single strand ends, which are particularly suitable for circularization, as well as smooth ends. Ligation for circularization of the molecules is preferably carried out in a sufficiently large dilution so that the probability that one end of the same cDNA molecule is closer to the other end than to any end of another molecule is sufficiently high. From the molecules that have been circularized in this way, constructs containing a TAG from each end of the molecule can then be obtained either enzymatically or by mechanical fragmentation. They will then have the following form:

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

[0097] By attaching two PBs to the ends of this construct, constructs of the form

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

or, depending on which 3' construct was used

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

which can be amplified starting from PB'' and PB''' and then sequenced, wherein PB'' and PB''' preferably but not necessarily have unequal sequences.

[0098] In order to obtain constructs of a defined length, size selection can be carried out, e.g. via HPLC or gel electrophoresis, or one or more tagging enzymes can be used.

[0099] For the use of type II enzymes, the 5'-end of the cDNA is provided with a construct of the form 5'-TES-(CAP site)-cDNA-3'

or

5'-PB-TES-(CAP site)-cDNA-3',

wherein the TES is preferably part of an adapter together with the PB.

3) Constructs for the quantification of cDNA that was not produced from poly-A-containing RNA

[0100] Unlike protein-coding eukaryotic mRNAs, many RNA molecules in eukaryotic cells and nearly all bacterial and viral RNAs do not contain a 3'-poly-A sequence. To quantify them, they are first ligated with a single- or double-stranded RNA oligonucleotide of known sequence, which can later serve as a primer binding site. Single-stranded oligonucleotides can be ligated specifically to the 5' or 3' side of a single-stranded nucleic acid with the aid of single-stranded RNA ligases, e.g. with the T4 RNA ligase, for example by protecting one end of the oligonucleotide from ligation while the 5' end can be phosphorylated. This allows an oligonucleotide containing a RAT and a PB (marker and adapter) to be ligated specifically to both ends of an RNA. For example, RATs can be ligated to non-protein-coding RNA, which can then be transcribed into DNA, amplified and sequenced. Corresponding constructs have the domain sequence:

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

or

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

wherein all molecules are single-stranded, or PBs can also exist as double-stranded RNA. As an alternative to ligation using RNA ligases, the RNA can be artificially provided with a poly-A strand at the 3' end of the RNA using a poly(A) polymerase.

[0101] After insertion of the poly-A sequence and conversion into double-stranded cDNA, RNAs with the same constructs as described above for mRNAs can also be quantified. However, since RNAs that do not code for proteins do not usually have a CAP structure, it is not possible to distinguish between full-length RNAs and broken molecules using procedural methods alone.

4) Constructs for quantifying genomic DNA fragments

4.1.) Constructs for quantifying genomic DNA regions: Chromatin-Immuno-Precipitation, Digital Karyotyping, Methylation-Specific Digital Karyotyping, Metagenomics

[0102] A method for the quantitative determination of DNA-protein binding is based on preparation of the sample by immuno-precipitation. Such methods are frequently used to determine the binding sites for transcription factors. For example, genomic DNA that has not been freed from bound proteins is enzymatically digested with a restriction enzyme or a non-specific cutting endonuclease or otherwise broken down (e.g. by ultrasound). The DNA/protein complexes are precipitated by immunoprecipitation with an antibody against a specific protein (e.g., a transcription factor). The co-precipitated DNA bound to the protein is freed from all proteins and converted into a form that can be quantified using RATs, with or without tagging enzymes. For this purpose, markers and, if necessary, adapters are ligated to the ends of the pre-extracted DNA fragments, the nucleic acid constructs contained therein are amplified and then sequenced. The result is the number of copies of a specific genomic DNA sequence, e.g. in a tissue sample, to which or in the vicinity of which a protein was bound.

[0103] Digital karyotyping (DK) enables the quantification of short DNA segments in the genome. The technique is used to examine chromosomal changes, duplications, and deletions and to detect the presence of foreign DNA. In this method, the DNA to be examined is first digested with a (methylation-sensitive) restriction enzyme. Now biotinylated markers or adapters can be ligated. The sample is then digested with a second, often cutting (methylation-sensitive) restriction enzyme and markers or adapters are ligated to the resulting interfaces. The resulting constructs must be labeled with at least one marker and can then be amplified and sequenced.

[0104] If markers or adapters are equipped with TES, a tagging enzyme can be used and, after binding to a solid phase, either the bound or the unbound marker TAG or adapter TAG combination can be purified and provided with a further marker or adapter, so that each marker with an adapter (or a marker) encloses a TAG. If markers and adapters are provided with a TES, in particular for EcoP15I, as is provided for in a preferred version, bound and unbound marker TAG and adapter TAG combinations can also be ligated together so that two TAGs of marker and/or adapter are enclosed.

[0105] Methylation-specific digital karyotyping (MSDK): This technique is used to examine the methylation status of genomic DNA. For this purpose, the first enzyme used in DK is replaced with a methylation-sensitive enzyme. This can only cut non-methylated recognition sequences. The remaining steps are the same as for the DK. Depending on the methylation status, different TAGs are produced from a single sample.

[0106] The following describes a particularly preferred variant of the DK, which includes the use of the tagging enzyme EcoP15I to produce nucleic acid constructs for the quantitative determination of genomic DNA according to the method of the invention.

[0107] For the analysis of genomic DNA fragments by DK, for metagenomics applications or after immunoprecipitation, a preferred method is to first digest the DNA with one or more restriction enzymes as described above. For the analysis of DNA from immunoprecipitated material, it can also be mechanically comminuted, e.g. by shearing or ultrasound.

[0108] The DNA fragments obtained are then ligated with a biotinylated oligonucleotide 1 (adapter or marker) which has a restriction enzyme cleavage sequence distal to the biotinylated end, wherein, after restriction digestion with the corresponding enzyme, a ligation site (LS1) with an overhanging single-stranded end is formed. For ligation with mechanically fragmented DNA, oligonucleotides (markers or adapters) with smooth ends are used. The ends of the fragmented DNA are also smoothed with the Klenow fragment before ligation. Furthermore, the adapter contains a primer binding site (PB1) at the biotinylated end. For EcoP15I to be used as a tagging enzyme, oligonucleotide 1 must also carry the sequence 5'-CAGCAG'-3' (TES) before the ligation site. After binding to a streptavidin-modified solid phase and washing

to remove unbound fragments, digestion is carried out with another enzyme. Restriction enzymes that have as many restriction interfaces as possible in the DNA to be analyzed, such as NlaIII, DpnII, TaqI, FaeI, Sau3A, MboI etc., are preferably used for this purpose. Another oligonucleotide (oligonucleotide 2) is ligated to the new restriction interface with the bound DNA. The oligonucleotide 2 contains a further recognition sequence for the restriction enzyme EcoP 15I (5'-CTGCTG-3' (TES)) followed by a restriction enzyme interface, from which a ligation site (LS2) emerges after restriction digestion. Ligation is performed using the DNA ligation sites created by digestion with the appropriate restriction enzyme, which often cuts the DNA. Oligonucleotide 2 is also protected from ligation at the end distal to the LS, e.g. by an amino-C7 modification.

[0109] The method steps described above result in the following product:

5'-CTGCTG-LS2-nucleic acid-LS1-CAGCAG-PB1-3'-(solid phase)

[0110] After digestion with the tagging enzyme EcoP15I, TAGs 25–27 bp in length are produced. These can now be ligated with another oligonucleotide 3. This oligonucleotide has the structure RAT-PB2, wherein ligation only takes place via the free end of the RAT, as the free end of the PB2 is protected from ligation, e.g. by an amino-C7 modification. After washing the solid phase, the bound nucleic acid constructs are detached from the solid phase, resulting in the following construct, which is particularly suitable for high-throughput sequencing:

PB2-RAT-TAG-LS1-CAGCAG-PB1.

[0111] Alternatively or additionally, RAT can also be integrated into oligonucleotide 1 containing PB1, resulting in the product PB2-TAG-LS1-CAGCAG-RAT-PB1 or PB2-RAT-TAG-LS1-CAGCAG-RAT-PB1.

[0112] After the second strand has been produced using a polymerase, a linker can be produced by digestion with an enzyme specific to the restriction interface, which has the base overhangs typical for the restriction interface and thus enables ligation with an interface with complementary base overhangs. Alternatively, ligation can also be performed with nucleic acid fragments with smooth interfaces. In particular, it is possible to link a RAT directly to a TAG without the interposition of a linker. In principle, the introduction of restriction interfaces via a left-hand sequence, which serves to generate overhanging interfaces, allows polynucleotides to be incorporated into the nucleic acid constructs in the desired orientation.

[0113] Degradation can take place mechanically, e.g. by means of shear forces, physically, e.g. by means of ultrasound or preferably by enzymatic degradation.

4.2.) Constructs for quantifying both ends of genomic DNA fragments using a circularization method

[0114] As already mentioned, it is often desirable to obtain one TAG from each end of a nucleic acid molecule of genomic DNA in order to quantify both ends. To fragment genomic DNA, it can be broken down into smaller fragments of any size using restriction enzymes, non-specific endonucleases or exonucleases, or by mechanical fragmentation. After binding these linear molecules to the aforementioned constructs of the form (1) TES-RAT-PB-RAT-TES or (2) TES-RAT-PB-PB'-RAT-TES, the molecules can be circularized and then amplified if the fragments do not exceed a certain size.

[0115] These reactions can take place in a solid phase or, preferably, in solution. Modification of one or both TES with a biotin group is a suitable method for binding to a solid phase.

[0116] The amplification of the circular molecule proceeds from the primer binding sites (PB), wherein in (1) the primers are at least partially complementary to each other and to one strand of the PB. (2) refers to two different primer binding sites, and the amplification initiated by the primers binding to these sites occurs in opposite directions. In both cases, circular molecules are formed that have the domain sequence

TAG-TES-RAT-PB-RAT-TAG

or

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

after cutting with the TE. It is often useful (e.g. for directed sequencing or cloning) to bind one or two further domains consisting of one or two adapter(s), one or two linker(s) or one or two primer binding site(s) to one or both ends of these constructs, so that constructs with the domain sequence shown in s) through v), for example, are then formed:

- 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'''

Here, the domains PB'' and PB''' can be replaced by the above-mentioned domains (adapters) in accordance with the invention.

Exemplary embodiments

Example 1: Use of RATs for the quantification of mRNAs.

[0117] The following describes an example of the use of RATs for the quantification of mRNAs in a preferred method. To do this, the mRNA is first transcribed into cDNA. Then, a TAG is obtained from practically every cDNA molecule, which is provided with a marker containing a RAT. The nucleic acid construct obtained is then amplified and sequenced.

cDNA synthesis with an oligo-dT primer:

[0118] Using a commercially available kit containing Trizol reagent (Invitrogen Inc.), approximately 1 mg of total RNA is extracted from a biological sample (an MVV leukemia cell line). 5 µg mRNA is isolated from the total RNA using a commercially available kit ("Oligtex Midi Kit," Qiagen N. V.). The cDNA is then produced with the "cDNA synthesis system", (Invitrogen Corp.), using the following oligonucleotide biotinylated at the 5' end, which contains the detection sequence for EcoP15I CAGCAG:

Biotin-5'-CTGATCTAGAGGTACCGGATCCCAGCAGTTTTTTTTTTTTTTTTTTT-3'

(SEQ ID No. 2;

[0119] The product is then converted to double-stranded DNA (cDNA synthesis system, Invitrogen Corp.) and dissolved in 20 µl LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA).

Digest with the anchoring enzyme DpnII:

[0120] The double-stranded cDNA (20 µl) is digested in a 200 µl reaction solution with 50 units of DpnII (New England BioLabs Inc, Ipswich, UK; NEB) in the supplied buffer at 37°C for 90 minutes. After digestion, the cDNA is extracted with TE-equilibrated phenol/chloroform/isoamyl alcohol (25:24:1, pH = 8.0), precipitated with ethanol, and dissolved in 100 µl LoTE buffer. 300 µl of streptavidin-coated magnetic particles (MP; Dynabeads M279, Dynal Biotech GmbH, Hamburg, DE) are washed with 200 µl of 1 × B&W solution (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl), using the Magnetic Particle Capturer (MPC; Promega GmbH, Mannheim, Germany) after adding the B&W solution to hold the magnetic particles to the vessel wall and replace the B&W solution. Next, add 100 µl of 2 × B&W solution (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) to the cDNA (100 µl) and replace the 1 × B&W solution of the MP with the cDNA solution. The MPs are then incubated with the biotinylated cDNA and the streptavidin MPs under rotation for 30 minutes at room temperature to allow binding. Using a magnet, the MPs to which the DpnII-digested cDNA is bound are washed twice with 200 µl of B&W solution and redissolved in 200 µl of LoTE.

Adapter ligation:

[0121] An adapter containing the recognition sequence for the tagging enzyme is ligated to the bound DNA. This is composed of the following single-stranded oligonucleotides A and B: Oligonucleotide B contains an amino-C7 modification at the 3' end to prevent unwanted ligation at this site and is phosphorylated at the 5' end to enable ligation on this side. The part of the 5'-end of the later double-stranded linker protected by the amino-C7 modification serves as a primer site for the subsequent pyrosequencing reaction using the Solexa method.

[0122] For this purpose, the two single strands of the oligonucleotide adapter

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

and

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

are dissolved in 0.2 mM Tris-EDTA at a concentration of 100 μ M and hybridized together to form a double strand by mixing both oligonucleotide solutions in a 1:1 ratio, heating to 96°C and then slowly cooling to room temperature. The linker contains the recognition sequence for the tagging enzyme EcoP15I (CAGCAG). This produces a double-stranded linker with a 5' GATC overhang that can be ligated with the DpnII restriction interface.

Ligase reaction:

[0123] The following ligation mixture is prepared:

6 μ l 5 $\times$ Ligase Buffer (Invitrogen, Corp.)

+5 μ l linker (100 μ M)

+17 μ l H₂O

= 28 μ l

[0124] The LoTE of the MP + cDNA mixture is replaced by the 28 μ l ligation mixture. The MP are mixed carefully, heated to 50°C for 2 minutes, and cooled to room temperature for 15 minutes. 2 μ l T4 ligase (5 U/ μ l, Invitrogen, Corp.) are added to the ligation mixture and the sample is incubated at 16°C for 2 hours, mixing the sample carefully every 20 minutes. After the ligation reaction, the unbound linkers are removed from the MP by washing three times with 300 μ l of 1 $\times$ B&W. The MPs are transferred to a new vessel and washed again once with 300 μ l of 1 $\times$ B&W and then twice with 1 $\times$ NEB4 buffer (New England Biolabs, Inc., NEB) and resuspended in 300 μ l of NEB4 buffer.

Cutting with the tagging enzyme:

[0125] The following reaction mixture is prepared:

10 μ l 10 $\times$ NEB3 buffer

10 μ l 10 mM ATP (NEB)

2 μ l EcoP15I (NEB)

1 μ l 100 $\times$ BSA (NEB) 77 μ l H₂O

= 100 μ l

[0126] The 300 μ l NEB3 buffer of the MP + cDNA + linker mixture is replaced with the 100 μ l reaction mixture. The sample is incubated at 37°C for one hour, mixing carefully several times. With the help of the MPC, the MP are held in place on the vessel wall while the supernatant can be removed and transferred to a new vessel. The supernatant contains the linker TAGs with a length of 52 bp (linker + TAG). The Linker-TAG fragments in the 100 μ l supernatant are extracted in a ratio of 1:1 with phenol/chloroform/isoamyl alcohol (25:24:1, pH = 8.0) solution, precipitated with 500 μ l ethanol + 67 μ l 7.5 M NH₄OAc + 2 μ l glycogen (NEB). After incubation at -20°C for 8 hours, the sample is centrifuged at 4°C and 13,000 g for 30 minutes, washed twice with 70% ethanol, dried, and dissolved in 20 μ l LoTE buffer.

Production of smooth-ended "blunting" of the adapter TAG fragments:

[0127] Cutting with EcoP15I produces adapter TAG fragments with protruding 5' ends. In order to later attach a RAT-containing marker with smooth ends, the ends are smoothed (blunting). The blunting reaction takes place in 50 μ l 1 $\times$ NEBuffer 2 (New England Biolabs, Inc.) with 33 μ M dNTP, 5 units polymerase I, large (Klenow) fragment (New England Biolabs Inc.), at 25°C and with an incubation time of 15 minutes. The sample is then filled up to 200 μ l with LoTE and extracted with phenol/chloroform/isoamyl alcohol (25:24:1,

pH = 8.0), precipitated with 1000 µl ethanol + 133 µl 7.5 M NH₄OAc + 2 µl glycogen (NEB) and dissolved in 10 µl LoTE.

Production and ligation with the RAT-containing marker:

[0128] The single-stranded marker with the following sequence:

phosphate 5' NNNNNNNNNNNTCGTATGCCGTCTTCTGCTTG 3' amino-C7
(=RAT-A), wherein (N)₈ represents the single-stranded RAT (SEQ ID No. 5)

is incubated with the primer with the following sequence

[0129]

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

using the Klenow fragment in the following solution for 30 minutes at 37°C, and thus produced double-stranded:

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] The double-stranded RAT-containing marker is purified using phenol-chloroform, precipitated with ethanol, washed, and after drying, taken up in 50 µl H₂O.

[0131] Ligation with the Adapter-TAG construct takes place under the following conditions:

10 µl TAG solution 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] The mixture is incubated at 16°C for 2 hours. The sample is then filled up to 200 µl with LoTE and extracted with phenol/chloroform/isoamyl alcohol (25:24:1, pH = 8.0) solution, precipitated with 1000 µl ethanol + 133 µl 7.5 M NH₄OAc + 2 µl glycogen (NEB) and dissolved in 10 µl LoTE.

PCR amplification:

[0133] First, PCR is performed with 25 cycles using various dilutions of the Linker-TAG-RAT ligation mixture to determine the concentration best suited for amplification. The PCR reaction mixtures, each 20 µl, are incubated for 2 minutes at 98°C for initial denaturation of the template DNA, followed by 25 cycles at 98°C for 15 seconds, followed by 15 seconds at 60°C and finally 15 seconds at 72°C. Phusion polymerase (Finzymes, Finland) is used as the polymerase.

The following primers were used

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

and

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

[0134] The PCR products are applied to an 8% polyacrylamide gel. After staining with ethidium bromide and visualization under UV light, a 79 bp marker adapter TAG band is visible in the gel. Next, for preparative

purposes, a 200 µl PCR reaction is performed using the dilution that yielded the best amplification.

Cleaning the Linker-TAG-RAT fragments:

[0135] The products of the preparative PCR are reapplied to an 8% polyacrylamide gel and the gel piece with the 79 base pair (bp) band is cut out with a scalpel. In a 1.5 ml Eppendorf tube, the gel piece is crushed with a pestle and incubated in 500 µl of an elution solution (0.5% NH₄OAc, 2 mM EDTA 0.1% SDS) overnight at 37°C. The eluted RAT-Linker-TAG construct is then precipitated from the elution solution with ethanol, washed twice with 70% ethanol, dried, and redissolved in 10 µl LoTE.

[0136] The concentration and quality of the PCR product is determined using gel electrophoresis. For testing, an aliquot of the purified RAT linker TAG construct is ligated into the pGEM®-T Easy vector (Promega), transformed into E. coli strain DH5alpha, white colonies are selected, and 5–10 RAT linker TAG constructs containing plasmids are sequenced using the Sanger method. Seven of the typical sequences obtained with different PB-RAT-TAG-LS-TES-PB' constructs are listed below. Each TAG is uniquely marked with a RAT. The RAT domains of the constructs are marked in bold, while the TAG domains are shown in italics.

Clone 1

CAAGCAGAAGACGGCATACG**ATCTCGACGGCCACATCCTCGGTATGACAGGAG**
ATCCTGCTGTAGAACTCTGAA (SEQ ID No. 9)

Clone 2

CAAGCAGAAGACGGCATACGAG**GTCAAGCGTGGGGGTGGGGGAGACACAATTT**
GATCCTGCTGTAGAACTCTGAA (SEQ ID No. 10)

Clone 3

CAAGCAGAAGACGGCATACGA**ACTTACAATTCTATGGTAGTACACACACACAG**
GATCCTGCTGTAGAACTCTGAA (SEQ ID No. 11)

Clone 4

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

Clone 5

CAAGCAGAAGACGGCATACGAC**CCATCCCCCAAGGCTGATATTTCCCTGGGA**
TCCTGCTGTAGAACTCTGAA (SEQ ID No. 13)

Clone 6

CAAGCAGAAGACGGCATACGAT**GTATTTGATTTATAATTTAATGGCTTGCAGATC**
CTGCTGTAGAACTCTGAACCTGT (SEQ ID No. 14)

Clone 7

CAAGCAGAAGACGGCATACGAT**GATGTAT**CCAATGGTTCCAGTCACATATTGAT
 CCTGCTGTAGAACTCTGAACCTGT (SEQ ID No. 15)

[0137] Using a BLAST search with the RAT-TAG construct in GenBank, the individual TAGs could be assigned to a specific human gene (Table 1):

Table 1: Tabular summary of the results of the BLAST search

Clone No.	Result of BLAST search
Clone 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 <pre> Query 9 GCCACATCCTCGGTATGACAGGAGATC 35 Sbjct 6328 GCCACATCCTCGGTATGACAGGAGATC 6354 </pre>

Clone 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 GTGGGGGTGGGGGAGACACAATTGATC 35 Sbjct 1285 GTGGGGGTGGGGGAGACACAATTGATC 1258
Clone 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
Clone 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
Clone 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 CCCAAGGCTGATATTTCCTGGGATC 34 Sbjct 48 CCCAAGGCTGATATTTCCTGGGATC 73

Clone 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
Clone 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

This is followed by a sequence listing in accordance with WIPO St. 25. This can be downloaded from the DPMA's official publication platform.

Claims

- Method for the quantitative determination of nucleic acids in a sample, **characterized in that**
 - nucleic acid constructs containing
 - at least one characteristic polynucleotide domain of a nucleic acid present in a sample that is to be quantitatively determined, and
 - at least one artificial polynucleotide domain with a random sequence and
 - at least one primer or polymerase binding site,
 - wherein a characteristic polynucleotide domain and at least one polynucleotide domain with a random sequence form a characteristic domain combination which is flanked by at least one primer or polymerase binding site,
 - are amplified in the presence of a polymerase, and
 - after amplification, at least the sequence of the polynucleotide domains with random sequence and a part of the characterizing polynucleotide domain are determined, wherein, for the quantitative determination of the nucleic acid present in the sample, sequence-identical or sequence-identical and sequence-like combinations of a characterizing polynucleotide domain and a polynucleotide domain with random sequence can be identified as copies of a characterizing polynucleotide domain of the nucleic acid originally present in the sample.
- Method for the quantitative determination of nucleic acids according to claim 1, **characterized in that** the amplification is a PCR amplification or an amplification by cloning.
- Method for the quantitative determination of nucleic acids according to one of claims 1 or 2, **characterized in that**
 - a sample containing a polynucleotide with a characteristic polynucleotide domain is mixed with a solution of a polynucleotide containing an artificial polynucleotide domain with a random sequence, wherein at least one of the two polynucleotides has a primer binding site,
 - ligation of at least one end of the polynucleotides containing the characteristic polynucleotide domains with the polynucleotide containing the polynucleotide domain with random sequence, so that the primer

binding site is not located between the characteristic polynucleotide domain and the polynucleotide domain with random sequence, and

c) the nucleic acid constructs obtained, marked with a polynucleotide domain with a random sequence, are amplified using a polymerase and at least the random sequence and the sequence of the characteristic polynucleotide domain are determined.

4. Method according to claim 3, **characterized in that** one end of the polynucleotides containing a characteristic polynucleotide domain and/or one end of the polynucleotides containing a polynucleotide domain with a random sequence are protected against ligation and/or enzymatic digestion, wherein, if a primer binding site is contained in the polynucleotide, this end is protected against ligation and/or enzymatic degradation.

5. Method according to claim 4, **characterized in that** the polynucleotides, which are protected at one end prior to restriction digestion and contain a characteristic polynucleotide domain, are digested with a restriction enzyme before the polynucleotides containing a polynucleotide domain with a random sequence are ligated to the free end of the remaining protected polynucleotide fragment.

6. Method according to one of claims 3 through 5, **characterized in that** the polynucleotides contain a binding site for binding a tagging enzyme and, after ligation of the polynucleotides containing the labeling polynucleotide domains with the polynucleotides containing a polynucleotide domain with a random sequence and after binding of the tagging enzyme, nucleic acid constructs are cleaved from the resulting ligated polynucleotides by the tagging enzyme, which contain a labeling polynucleotide domain and which contain a polynucleotide domain with a random sequence.

7. Method according to claim 6, **characterized in that** the tagging enzyme used is a type II or type III restriction enzyme.

8. Method according to one of claims 6 or 7, **characterized in that** the tagging enzyme used is EcoP15I, MmeI, or BsmFI.

9. Method according to claim 1, **characterized in that** the nucleic acid construct used has one of the following domain sequences:

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; with n = 2 to 50, preferably with n = 2 to 10, PB-RAT-(TAG)_n-RAT-PB; with n = 2 to 50, preferably with n = 2 to 10,

PB-RAT-(TAG)_n; with n = 2 to 50, preferably with n = 2 to 10,

PB-RAT-TAG-RAT-TAG,

PB-TAG-RAT-TAG-RAT,

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

PB-(RAT-TAG)_n-PB'; with n = 3 to 50, preferably with n = 3 to 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';

with n = 0 to 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;

with m = 0 or 1 independently of each other for each m, wherein at least one m = 1,

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';

with n = 3 to 50, preferably with n = 3 to 10 and m = 0 or 1 independently of each other for each m,

PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''; with m = 0 or 1 for each m independently of one another, wherein at least one m = 1,

PB-RAT-(TES)_m-TAG-(TES)_m-PB'-PB''-RAT-(TES)_m-TAG-(TES)_m-PB'''-PB''''-RAT-(TES)_m-TAG-(TES)_m-PB'''''; with m = 0 or 1 for each m independently of one another, wherein at least one m = 1,

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' with n = 0 to 10 and with m = 0 or 1 for each m independently of one another, wherein at least one m = 1,

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 or

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

PB, PB', PB'', PB''', PB''''', PB'''''' primer binding sites may have the same or different sequences, RAT stands for a polynucleotide domain with a random sequence, TAG stands for a polynucleotide domain with a characteristic sequence, and TES stands for a tagging enzyme binding site.

10. Method according to claim 1, **characterized in that** the nucleic acid construct used has one of the following domain sequences:

cDNA oligo A/T RAT promoter,

cDNA oligo A/T RAT PB promoter,

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

5'-promoter-RAT-mRNA-3',

5'-biotin-PB/promoter-RAT-TES mRNA,

5'-biotin-PB/promoter-RAT-TES-cDNA-oligo-A/T-3',

5'-biotin-PB/promoter-RAT-TES-TAG-3',

5'-biotin-PB/promoter-RAT-TAG,

5'-biotin-PB/promoter-TAG-TES-PB-3',

5'-biotin-PB/promoter-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, or
TAG-TES-RAT-PB-PB'-RAT-TAG.

11. Marker ensemble for carrying out the method according to one of claims 1 through 10 for the quantitative determination of nucleic acids, **characterized in that** the markers contained therein each comprise an artificial polynucleotide domain with random sequences that are different from one another and a primer binding site.

12. Marker ensemble according to claim 11, **characterized in that** the markers contain a binding site for a tagging enzyme, wherein the tagging enzyme binding site is located at the distal end relative to the primer binding site of the marker.

13. Use of an ensemble of markers according to one of claims 11 or 12 for the quantitative determination of nucleic acids in a sample.

14. Use of a method according to one of claims 1 through 8 for the quantitative determination of mRNA, cDNA, aptamer DNA or genomic DNA, for the methylation-specific quantitative determination of genomic DNA, for the quantitative determination of protein-bound genomic DNA.

This is followed by 4 pages of drawings.

1) AGCTGTCAGCCT CTGATGGTGTTGCTAGGTACTGAATT
RAT TAG

2) CATGCTAGTCGG CTGATGGTGTTGCTAGGTACTGAATT
RAT TAG

3) GGCTTATCGTAG CTGATGGTGTTGCTAGGTACTGAATT
RAT TAG

4) GGCTTATCGTAG CTGATGGTGTTGCTAGGTACTGAATT
RAT TAG

Fig. 2 (continued)

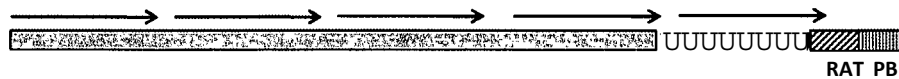
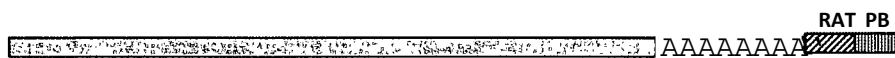
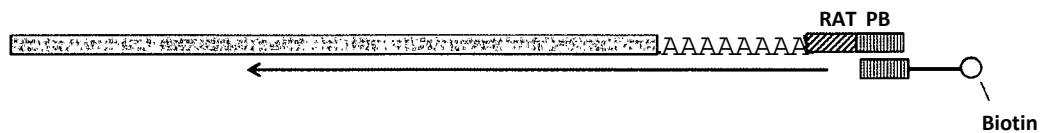
1.4) First-stage synthesis: Reverse transcription with oligohexamers**1.5) Amplified antisense DNA****1.6) Second-strand synthesis using a (optional) biotinylated primer complementary to PB**

Fig. 2 (continued)

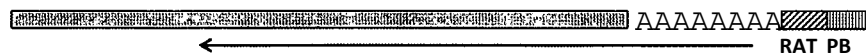
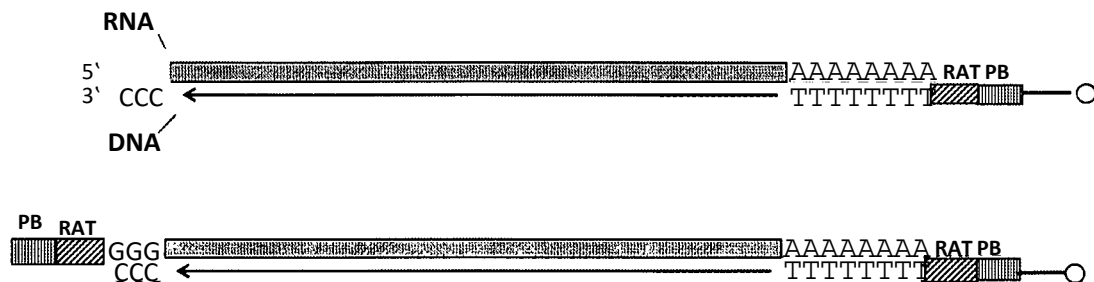
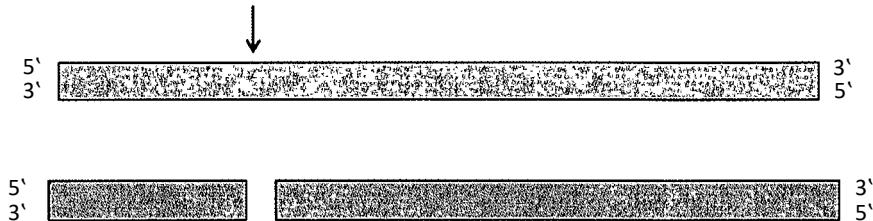
2) Analysis of mRNA without prior amplification**3) Analysis of mRNA without prior amplification using TES insertion****4) Amplification of mRNA with incorporation of markers at one or both ends**

Fig. 3

1. Digestion with methylation-insensitive restriction enzyme



2. Ligation of a first marker

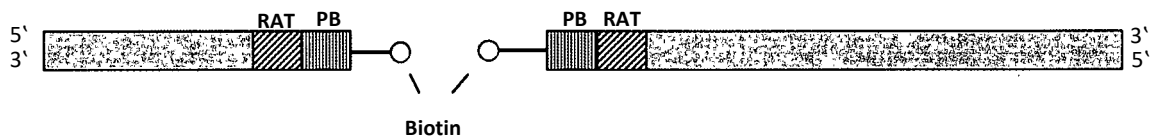
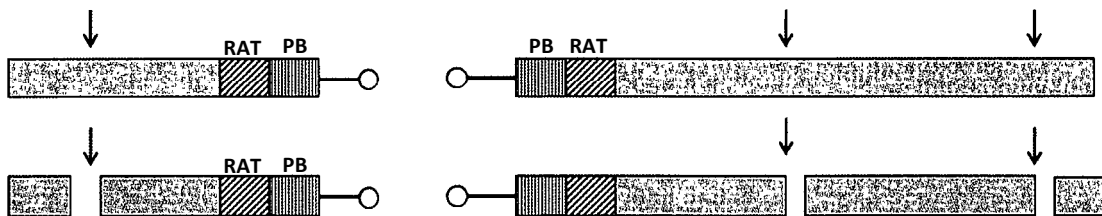
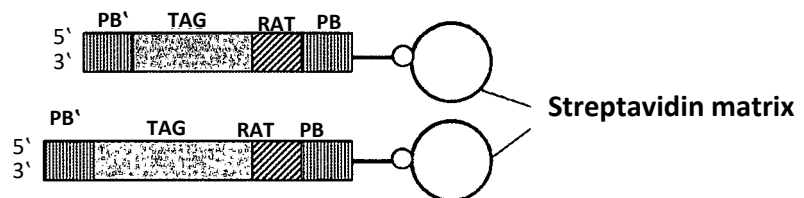


Fig. 3 (continued)

3. Digest with a second enzyme



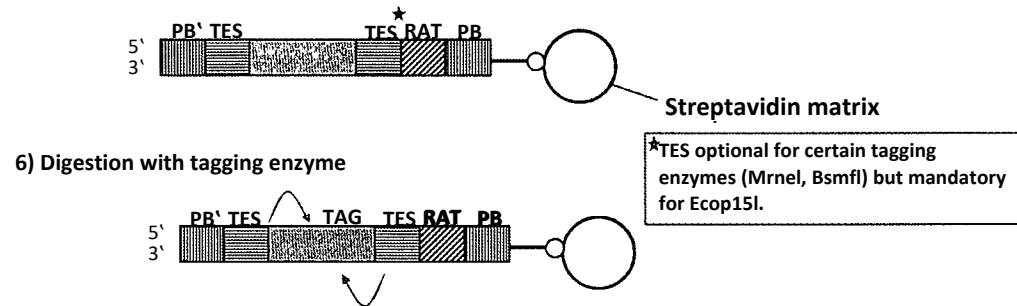
4. Binding to streptavidin matrix, ligation with adapter containing PB'



5. Sequencing

Fig. 4

5) Nucleic acid constructs, constructed with markers and adapters, which contain a tagging enzyme binding site (TES) distal to PB or PB', are obtained in place of step 4) from Fig. 3.



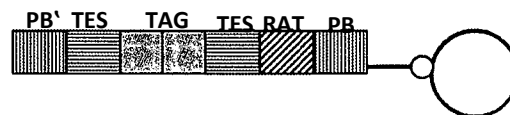
7.1) Use of the unattached end as a tag:



7.2) Using the tied end as a tag:

**Fig. 4 (continued)**

7.3) Using both ends as tags:



8.) Sequencing