

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2016/207379 A1

(43) International Publication Date  
29 December 2016 (29.12.2016)

WIPO | PCT

(51) International Patent Classification:  
*C12Q 1/68* (2006.01)

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(21) International Application Number:  
PCT/EP2016/064719

(22) International Filing Date:  
24 June 2016 (24.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
62/185,111 26 June 2015 (26.06.2015) US  
PA 2015 70449 7 July 2015 (07.07.2015) DK

(71) Applicant: SAMPLIX S.A.R.L. [LU/LU]; 58, Rue Charles Martel, 2134 Luxembourg (LU).

(72) Inventors: MIKKELSEN, Marie Just; Stenmaglevej 13, 2700 Brønshøj (DK). KVIST, Thomas; Dalengen 11, 4000 Roskilde (DK). MADSEN, Esben Bjørn; Jagtvej 33, st. tv., 2200 Copenhagen N (DK).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

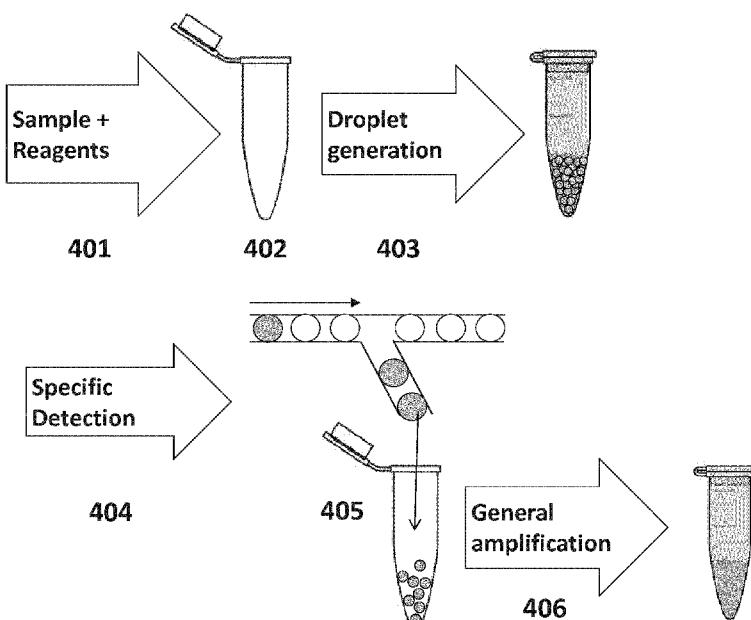
(74) Agent: GUARDIAN IP CONSULTING I/S; Diplomvej, Building 381, 2800 Kgs. Lyngby (DK).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: TARGETED ENRICHMENT OF LONG NUCLEOTIDE SEQUENCES USING MICROFLUIDIC PARTITIONING

Figure 4



(57) Abstract: The invention relates to a method for enrichment or isolation of a long nucleotide fragment comprising a known nucleotide sequence element, i.e. a sequence encoding a conserved active site or domain, the method being applicable i.a. to high throughput screening for DNA fragments containing a known sequence element.

**Title: Targeted enrichment of long nucleotide sequences using microfluidic partitioning**

**TECHNICAL FIELD**

The invention relates to a method for enrichment or isolation of a long nucleotide 5 fragment comprising a known nucleotide sequence element, i.e. a sequence encoding a conserved active site or domain, the method being applicable i.a. to high throughput screening for DNA fragments containing a known sequence element.

**BACKGROUND OF THE INVENTION**

10 Sequencing of DNA is a major driver in genetics research. The 'next generation sequencing' technological revolution is gathering momentum as new robust high-throughput sequencing instruments are becoming available. New and improved methods and protocols have been developed to support a diverse range of applications, including analysis of genetic variation. As part of this, methods 15 have been developed that aim to achieve targeted enrichment of genome sub-regions such as targeted cancer panels or complete human exomes. By selective recovery of genomic loci of interest, costs and effort can be reduced significantly compared with whole-genome sequencing.

20 Current techniques for targeted enrichment fall into three categories; Hybrid capture, selective circularization, and PCR amplification. In hybrid capture techniques, short fragment libraries (typically 100-250 base pairs) are hybridized specifically to complementary DNA fragments so that one can physically capture and isolate the sequences of interest. Selective circularization encompasses methods wherein single-stranded DNA circles including target sequences are 25 formed, creating structures with common DNA elements that are then used for selective amplification of the target sequence. Finally, PCR amplification based enrichment is directed toward the target region by conducting multiple PCR reactions in parallel.

30 Common for the current enrichment methods is that they require a significant knowledge of the target sequence, a relatively pure sample, a significant amount of target sequence, and that they produce relatively short sequences.

Accordingly there is a need for methods that can be used to analyze complex DNA samples, for example mammalian genomic DNA samples, where the frequency of the target is low (i.e. 1 target copy per genome), and where the analysis does not depend on the individualized analysis of each DNA molecule in 5 the sample, which is both expensive and time consuming.

## **SUMMARY OF THE INVENTION**

The invention, according to a first embodiment, provides an *in vitro* method for enriching for one of more target DNA molecule from a sample of mixed DNA 10 molecules comprising the steps of:

- a) providing a liquid sample of mixed DNA molecules comprising one or more specific target DNA molecule and reagents for specific detection of at least one of said target DNA molecules (401),
- b) formation of a multiple of liquid droplets from said liquid sample 15 (403),
- c) specific detection of droplets containing at least one of said target DNA molecules (404), wherein each droplet contains less than 0.5, preferably less than 0.25 or even more preferably less than 0.1 of said one of more target DNA molecule on average (404), and
- d) physically selecting droplets containing at least one of said target DNA 20 molecules (405),
- e) general amplification of the DNA molecules in the coalesced selected droplets, where said DNA molecules are in a total amount of less than 300fg, wherein general amplification reagents are added to the coalesced droplets to form a general amplification reaction mixture, and wherein from at least  $1.2 \times 10^6$  and up to a maximum of  $1.2 \times 10^9$  25 droplets are formed for each 5 $\mu$ l of the reaction mixture.

The invention, according to a second embodiment, provides an *in vitro* method for enriching for one of more target DNA molecule from a sample 30 of mixed DNA molecules comprising the steps of:

- a) providing a liquid sample of mixed DNA molecules comprising one or more specific target DNA molecule and reagents for specific detection of at least one of said target DNA molecules (401),
- 5 b) formation of a multiple of liquid droplets each comprising mixed DNA molecules from said liquid sample (403),
- c) specific detection of droplets containing at least one of said target DNA molecules (404), wherein each droplet contains less than 0.5, preferably less than 0.25 or even more preferably less than 0.1 of said one of more target DNA molecule on average (404), and
- 10 d) physically selecting droplets containing at least one of said target DNA molecules (405), wherein the frequency of the target DNA molecule compared to its frequency in the sample of mixed DNA molecules in step (a) is increased between  $0.01 \times (\text{total number of droplets containing DNA molecules}) \times (\text{number of droplets with target DNA})^{-1}$  and  $100 \times (\text{total number of droplets containing DNA molecules}) \times (\text{number of droplets with target DNA})^{-1}$ ,
- 15 e) general amplification of the mixed DNA molecules in the coalesced selected droplets, where the DNA molecules are in a total amount of less than 300fg, wherein general amplification reagents are added to the coalesced droplets to form a general amplification reaction mixture, and wherein from at least  $1.2 \times 10^6$  and up to a maximum of  $1.2 \times 10^9$  droplets are formed for each  $5\mu\text{l}$  of the reaction mixture.

20 In an alternative to the first or second embodiments, the total amount of DNA (originating from the selected droplets) in step e) is less than 200fg, or less than 150fg, or less than 100fg, or less than 80fg, or less than 60fg, or less than 40fg, or less than 20fg, or less than 10fg, or less than 5fg, or less than 3fg, or less than 2fg, or less than 1fg.

30 **LEGENDS TO THE FIGURES**

**Figure 1:** Comparison of PINSSwift to current methods of specific DNA enrichment. Upper part: The filled black lines represent a DNA target of

approximately 30 kb. The dotted lines represent the nucleotide sequence information in respect of the target sequence that is needed in advance in order to perform the enrichment method. The nucleotide sequence information required for PINSSwift can be located at any position on the DNA target sequence. Lower part (below the arrow): The black lines represent enriched DNA target molecule showing the difference in the length of the target DNA molecules after the three different types of enrichment.

**Figure 2:** Enrichment of target DNA molecule (fold enrichment relative to the DNA in the starting sample based on the number of target molecules per ng of total DNA) after 1, 2, 3, and 4 rounds of PINSSwift shown for 1, 5 and 20 targets multiplexed in one reaction assuming four positive droplets per target sequence and a total of 20,000 droplets.

**Figure 3:** Correlation between the number of target DNA molecules, the average number of positive droplets and the resulting enrichment of target DNA molecules; exemplified using one round of PINSSwift, 20,000 droplets and 10 positive droplets per target.

**Figure 4:** Outline of a scheme for performing PINSSwift

401: DNA sample containing one or more DNA target and reagents for the detection reaction are mixed.  
20 402: Components required for droplet generation are added to the mixture.  
403: Droplets containing the DNA and reagents are generated.  
404: Reaction steps such as PCR cycling needed for specific detection (optional) are performed.  
405: Positive droplets containing target DNA are detected, selected and sorted  
25 using an apparatus for physical selection of droplets.  
406: General amplification of the DNA from the coalesced positive droplets.

**Figure 5:** Outline of a scheme for performing PINSSwift as described in Example 1.

30 501: A DNA sample containing the DNA target and reagents for PCR are mixed.  
502: Components required for droplet generation are added to the mixture.  
503: Droplets containing the DNA and reagents are generated.

504: PCR amplification of a fragment of the target nucleotide.

505: Positive droplets containing target DNA are detected based on fluorescence generated in PCR, selected and sorted in a microfluidic chip.

506-508: The collected, sorted positive droplets are coalesced, reagents for

5 Phi29 amplification are added to the water phase, new droplets are formed from the water phase by adding oil and mixing, and the mix is incubated at 30°C

509: The droplets are coalesced and the amplified DNA is purified and sequenced.

10 **Figure 6:** Outline of a scheme for performing PINSSwift as described in example 2.

601: A DNA sample containing the DNA target and reagents for PCR are mixed.

602: Components required for droplet generation are added to the mixture.

603: Droplets containing the DNA sample and reagents are generated.

15 604: PCR amplification of a portion of the target DNA molecule.

605: Positive droplets containing target DNA molecule are detected based on fluorescence generated in PCR, selected and sorted in a microfluidic chip.

606-608: The collected, sorted positive droplets are coalesced, reagents for Phi29 amplification are added to the water phase, new droplets are generated

20 from the water phase and the droplet emulsion is incubated at 30°C.

609: The sample is further amplified resulting in Sample A (610).

610: A part of Sample 608 is used for a second round of enrichment by dilution and repetition of steps 601 to 606.

611-614: The sorted droplets (612) are coalesced, reagents for Phi29

25 amplification are added and the DNA fragments are amplified in emulsion to generate Sample B (614).

**Figure 7:** Outline of a scheme for performing PINSSwift as described in example 3.

30 701: A DNA sample containing the DNA target and reagents for PCR are mixed.

702: Components required for droplet generation are added to the mixture.

703: Droplets containing the DNA sample and reagents are generated.

704: PCR amplification of a portion of the target DNA molecule using dUTP in the reaction mixture.

705: Positive droplets containing the target DNA molecule are detected based on fluorescence generated in PCR, selected and sorted in a microfluidic chip.

706: The collected, sorted positive droplets are coalesced, UDG is added and the reaction is incubated at 37°C to remove the PCR product used for detection and

5 then the reaction is terminated by incubation at 95°C resulting in sample 707.

708-709: Reagents for Phi polymerase amplification are added to the water phase, emulsion oil is added, the sample is vortexed and the emulsion is incubated at 30°C, and the droplets are coalesced to generate sample 710.

10 **Figure 8:** An agarose gel (stained for DNA) comprising the size-separated products of Phi29 amplification of an *E. coli* DNA template. The Phi polymerase amplification reaction mixture with or without template was prepared, and the reaction was allowed to proceed; or the reaction mixture was converted to droplets and then the reaction was allowed to proceed; 1) 1 ng/µL *E. coli* DNA template, no droplets, 2) No DNA template, no droplets, 3) 1 ng/µL *E. coli* DNA template, reaction in droplets, 4) No DNA template, reaction in droplets, 5) 1 ng/µL *E. coli* DNA template, no droplets, no Phi29 polymerase. Marker: 1 kb+ DNA ladder as size marker.

15

20 **Figure 9:** Quantitative PCR specific detection and quantitation of the *E. coli* DNA produced by Phi polymerase amplification of an *E. coli* DNA template (analyzed samples correspond to the products of Phi polymerase amplification size separated on the agarose gel shown in Figure 8. The templates used for qPCR were the products of: 1) 1 ng/µL *E. coli* DNA template amplified by Phi

25 polymerase, no droplets (B01-B02), 2) No template control amplified by Phi polymerase, no droplets (B03-B04), 3) 1 ng/µL *E. coli* DNA template amplified by Phi polymerase in droplets (B05-B06), 4) No template control amplified by Phi polymerase in droplets (B07-B08), 5) Positive *E. coli* DNA control (B09-10).

30 **DETAILED DESCRIPTION OF THE INVENTION**

Detection and characterisation of a target DNA fragment (or molecule) in a complex mixture of DNA fragments (or molecules), particularly a low copy number target molecule, requires a sensitive detection method combined with methods for purification and analysis of the target DNA. It is known in the art,

that the most sensitive PCR reactions are obtained when the DNA fragment is short, such as 100-250 base pairs. Long range PCR, designed to amplify DNA fragments that are longer than 250 base pairs, for example 500 to 5000 base pairs in length, are not capable of detecting low amounts of target DNA fragment

5 present in a sample of mixed DNA fragments containing high amounts of background DNA. Similarly, hybridization based methods require the sample to be relatively pure to avoid non-specific hybridization. The best described ways of obtaining sequence information from a sample of mixed DNA fragments or a very dilute sample has been based on sequencing the entire mixture of DNA

10 fragments by e.g. next generation sequencing methods, and, although the cost of next generation sequencing is rapidly decreasing, the cost of sequencing thousands of genomes is still high. However, even this approach has its limitations, since the PCR amplified DNA fragments themselves yield short sequences (100-250 base pairs) that must be further assembled to determine

15 translocations, insertions or deletions or polymorphisms that may be detected over long range regions of DNA fragments of previously unknown sequence.

The *in vitro* method for enriching for one or more target DNA molecule from a sample of mixed DNA molecules according to the present invention, employs DNA detection methods (such as PCR) solely for the purpose of identifying DNA

20 molecules (fragments) that comprise the one or more target DNA molecule. Since short DNA sequences are sufficient to identify the target DNA, the present method takes maximum advantage of the most sensitive forms of PCR for detection of a DNA target molecule in a complex DNA mixture. The detection reaction is performed in droplets prepared from the sample of mixed DNA

25 molecule combined with detection reagents. This has the advantage that the reaction conditions and reagents are optimised for the detection step, which is particularly advantageous for PCR-based detection with respect to Mg<sup>++</sup> ion concentration (which is 10 fold lower than the optimal Mg<sup>++</sup> ion concentration for Phi polymerase general amplification), and target-specific primers. Since the

30 droplet size is typically small (1 – 100 $\mu$ m in diameter), the total volume of the combined droplets comprising target DNA will be small relative to the subsequent amplification reactions, allowing an optimisation of reaction conditions for the subsequent steps of the method.

The characterisation of one or more target DNA molecule detected in a sample of mixed DNA molecules requires a sufficient number of copies of the detected molecule. The enzyme Phi29 polymerase, derived from the bacteriophage  $\Phi 29$ , can be used for whole genome amplification (WGA). Using a method of multiple displacement amplification (MDA), fragments tens of kilobases in length can be produced using the Phi29 polymerase together with non-specific hexameric primers that anneal at intervals along a DNA template. However, it is known that Phi29 polymerase requires a certain amount of input DNA to function in an amplification reaction. Main suppliers of Phi29 polymerase such as Qiagen and Fidelity systems state in their product specifications that the minimum input for the Phi polymerase reaction is 1 ng or higher ([www.qiagen.com/dk/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/repli-g-single-cell-kit](http://www.qiagen.com/dk/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/repli-g-single-cell-kit), [www.fidelitysystems.com/phi29\\_hexamers.html](http://www.fidelitysystems.com/phi29_hexamers.html)).

Additionally, locus underrepresentation and background products have been reported when low amounts of DNA input material are used. The method according to the invention requires a polymerase that is capable of amplifying  $\leq$ less than 300fg of DNA (corresponding to 5565 linear DNA molecule of 50,000 nucleotides). This is more than 3000 fold lower than the reported template requirements of the Phi29 polymerase. Surprisingly, the Phi29 polymerase was found to provide an effective amplification of detected and selected DNA molecules using the method of the invention, even when the input DNA was below 5 fg of DNA. As illustrated in Example 1, as little as 1.7fg of DNA (corresponding to 1.3 linear DNA molecules of 50,000 base) were effectively amplified using Phi29 polymerase. Thus, not only was Phi29 polymerase unexpectedly found to amplify such low amounts of DNA molecules but additionally it was unexpectedly found that the general amplification reaction can be performed in a very small volume, such as droplet (1 – 100 $\mu$ m in diameter). Where the general amplification volume is performed in such droplets, the degree of bias is further reduced, since the production of non-specific amplification products (such as those derived from amplification of non-specific hexameric primers and/or contaminating DNA not originating from the sample of mixed DNA molecules) is minimised. Surprisingly, the present invention reveals that amplification of contaminating DNA by Phi29 polymerase in a sample can be

eliminated by performing the general amplification reaction of DNA molecules in droplets, in particular droplets having a volume of about 5pL (see Example 4).

The present invention pertains to an *in vitro* method in which the concentration of a specific target DNA molecule is increased relative to the concentration of 5 total DNA in a sample, by 1) dilution of a sample into multiple sub-compartments such as droplets (separation) containing reagents for detection of specific target, 2) detection of the specific target sequence within the droplets and, 3) physical selection of droplets containing the target sequence (selection), 4) addition of chemicals for general DNA amplification to the selected DNA 10 containing the target sequence, 5) general amplification of the selected DNA containing the target DNA molecule.

The invention is based on the principle that, if only a fraction of the droplets contain the target DNA molecule, the concentration of the target DNA molecule relative to total DNA is higher in these droplets, compared to the concentration 15 in the original sample. The fraction of droplets containing the target determines the degree of enrichment; if the fraction is low, the enrichment is high (Figure 3).

In the context of the present invention, the presence or absence of the target 20 DNA molecule in a sample or a dilution thereof or droplet, is defined by the presence of detectable target molecule in the sample, or a dilution thereof, or droplet using the selected method of detection (e.g. PCR).

The target can be further enriched by further rounds of selection (as above), until it can be sequenced by standard methods such as Sanger sequencing, sequencing by synthesis, Pyrosequencing or similar detection of DNA sequence 25 or by PCR, hybridization or other detection assays, or it can be used directly from the first round of selection.

Droplets amplified and sorted according to the invention contain DNA fragments of 5-100 kb containing the sequence used for identification.

Surprisingly little prior sequence information is needed for the enrichment 30 according to the invention. In comparison to current enrichment technologies, only approximately 40 nucleotide base pairs of specific target information is required as compared to at least 5000-8000 and 300 base pairs respectively for

hybridization based and long range PCR based methods respectively (Figure 1). The small amount of sequence information needed in combination with the large DNA fragments obtained from enrichment makes it possible to sequence into regions of a given DNA fragment that are only partially known. These regions 5 can contain unknown translocations, insertions or deletions or they can be regions comprising polymorphisms.

The method of the invention is surprisingly efficient, whereby the extent of DNA sequencing can be reduced by a factor of more than 1 billion for 5 multiplexed DNA targets by three rounds of PINSSwift or a factor of 25 million for one target 10 sequence using two rounds of PINSSwift (Figure 2).

### **I: Partitioning Isolation of Nucleotide Sequences (PINSSwift)**

The essential steps of the PINSSwift method are further described below:

- a) Providing a DNA sample potentially comprising one or more specific target DNA molecule and reagents for detecting the presence of the target molecule (401)  
15 A sample of mixed DNA molecules potentially comprising a target DNA molecule, is selected for performing PINSSwift. One or more nucleotide sequences of at least 10 (or 15, more preferably 40) unique nucleotides located within the target DNA molecule is selected for screening and 20 detecting the DNA molecule by a desired method, such as PCR detection, DNA detection with hybridization probes or similar. The target DNA molecule is typically larger than 500 base pairs, such as between 2000 and  $10^8$  base pairs, preferably between 2,000 and 200,000 base pairs, more preferably between 10,000 – 100,000 base pairs. A target DNA 25 molecule may contain more than one unique nucleotide sequence, each sequence corresponding to a given genetic marker, for example one genetic marker sequence diagnostic of an infectious agent and a second marker diagnostic of an antibiotic resistance gene. Typically, the frequency of the target DNA molecule in the sample of mixed DNA 30 molecules is less than  $10^{-2}$ , it may for example lie between  $10^{-3}$  and  $10^{-9}$ , or for example between  $10^{-4}$  -  $10^{-7}$ , or even less than  $10^{-11}$  (calculated as base pairs of target sequence divided by base pairs of total DNA in the sample). Prior to amplification, the liquid sample of mixed DNA molecules

is serially diluted by a desired number of dilutions until each droplet in the subsequent droplet formation step contain less than 0.01 target DNA molecule on average, preferably less than 0.001 or even more preferably less than 0.0001 specific target DNA molecule on average. Typically, the 5 total number of droplets formed in step b) of the method is from between  $5 \times 10^3$  and  $1 \times 10^{10}$ , or from between  $2 \times 10^4$  and  $1 \times 10^7$ , or between  $5 \times 10^5$  and  $5 \times 10^6$ . Thus, if the liquid sample of mixed DNA molecules is 10 separated into 100,000 droplets, on average the target DNA molecule will be present in less than 1,000 of these droplets, preferably less than 100 of these droplets, even more preferably less than 10 of these droplets. The presence or absence of the target DNA molecule in a droplet is 15 defined herein as the presence or absence of detectable target DNA molecule when employing methods for specific detection of the target DNA molecule, such as those exemplified in the present application. This dilution is performed to ensure target enrichment; if the average number 20 of droplets containing target is low, the frequency of target relative to non-target molecules within the droplet is high. The frequency and abundance of the target DNA molecule in the mixed DNA sample may be determined by DNA concentration measurement, PCR, real time PCR, by hybridization based assays or by assays detecting an RNA or protein 25 product of the target sequence. Reagents for PCR detection include a PCR polymerase (e.g Taq polymerase), a  $Mg^{++}$  concentration of 1.5–2.5 mM in 1X Standard Taq Reaction Buffer, and 200  $\mu M$  dNTPs of each deoxynucleotide, that is typically optimal for most PCR products generated with Taq DNA Polymerase. Oligonucleotide primers, generally 30 20–40 nucleotides in length, whose sequence is capable of specifically binding to the 5' and 3' ends of the known sequence of the target DNA molecule.

b) Formation of a multiple of droplets containing said DNA sample each containing less than 0.5, preferably less than 0.25 or even more preferably less than 0.1 specific target DNA molecule on average (403).

Droplets containing diluted sample DNA and reagents necessary for 35 specific detection of target DNA are generated using any method of droplet generation to isolate target DNA sequences in closed compartments. Suitable methods for droplet generation include active

methods such as acoustic energy ejected droplets, dielectrophoresis (DEP) and electrowetting on dielectric (EWOD), and passive methods such as shear focusing [1], or T-junction, x-junction and flow focusing [2], vortexing, shaking, and sonication. In addition to droplets, the general amplification can occur in other micro-volume compartments, such as reaction chambers in microfluidic chips.

5 c) Specific detection of droplets containing said specific target DNA (404)

Following droplet generation in step b) the droplets are screened for the presence of the target DNA molecule using the desired detection technique. In at least one or more screened droplets that are shown to contain the target DNA molecule, the frequency of the target DNA molecule will be increased compared to its frequency in the sample of mixed DNA molecules in step (a) According to a first embodiment of the invention, the average total number of DNA fragments per droplet (e.g. single target in a sample, such as one gene from a human genome): is from 1 to 1,000, or 2 to 100, or 3 to 10 DNA fragments. Where there is on average only one DNA fragment per droplet, then the positive droplets will on average comprise the target DNA molecule in pure form. According to the second embodiment of the method, all droplets formed in step b) of the method comprise mixed DNA and the increase in frequency, obtained by selection of the positive droplets is typically between  $0.1 \times (\text{total number of DNA-containing droplets}) \times (\text{number of droplets with target DNA})^{-1}$  and  $10 \times (\text{total number of DNA-containing droplets}) \times (\text{number of droplets with target DNA})^{-1}$ .

10 20 25 The number of droplets containing target is typically between 5 and 1000 per target sequence. The total number of droplets is at least 5,000, but typically greater than 20,000 and often greater than 500,000.

30 The presence of the target DNA molecule in the droplets may be determined by PCR including qPCR, by hybridization based assays or by assays detecting an RNA or protein product of the target sequence. The reagents for specific detection may contain dUTP to make it possible to selectively inactivate the DNA amplified in the detection step using UDG at a subsequent step.

d) Physical selection of droplets containing said specific target DNA (406)

Based on the detection of target DNA in step c) droplets are sorted into at least two different streams. When more than one specific target is detected in step c), the droplets may be sorted into 3, 4, 5 or more streams. In the stream containing droplets wherein the target is detected, the target is enriched as compared to the sample of mixed DNA molecules in step a).

5 e) General amplification of DNA molecules from selected droplets containing specific target DNA molecules (406)

DNA from selected droplets containing specific target DNA, and wherein total amount of DNA is less than 300fg, is amplified using any method of total DNA amplification to increase the abundance of the DNA in each sample. A suitable general amplification method is Multiple Displacement Amplification (MDA)[4]. In a preferred embodiment the enzyme employed for performing the general amplification is Phi29 polymerase; preferably the general amplification enzyme consists of Phi29 polymerase, whereby general amplification does not require additional polymerases. It may be necessary to add reagents for general amplification by breaking, mixing, or fusing droplets or to break emulsion by applying force such as centrifugation and/or sonication.

10 20 In one embodiment, droplets, generated from a liquid sample comprising the selected coalesced droplets containing specific target DNA and the reagents for general amplification, are prepared using any method of droplet generation to isolate target DNA sequences in closed compartments (see above). General amplification then takes place in the

15 25 Preferably the general amplification reaction is performed in droplets that are at least 1  $\mu$ m in diameter on average (corresponding to an average droplet volume of at least 0.00052 pL; allowing the formation of  $\leq$ 1.9 billion droplets/ $\mu$ L general amplification reaction mixture). Preferably the droplets have a diameter

20 30 of from 2 to 20  $\mu$ m on average (corresponding to an average droplet volume of 0.0042 pL; with the formation of 240 million droplets/ $\mu$ L; up to an average droplet volume of 4.2 pL; with the formation of 240,000 droplets/ $\mu$ L). Accordingly, where the volume of the general amplification reaction is at least 5 $\mu$ L, then the general amplification reaction will be performed in at least  $1.2 \times 10^6$  droplets; and up to a maximum of  $1.2 \times 10^9$

droplets for each 5 $\mu$ l of reaction mixture. Following amplification, all droplets are then coalesced, and the target DNA molecules may then be further detected, selected and amplified in a subsequent cycle of steps a) to e).

5

Optional steps:

f) Inactivating, degrading or removing DNA produced for specific detection of target DNA

When the enriched target DNA molecule is used for further rounds of PINSSwift or other applications where the presence of the detection product interferes with these further processing, the amplification of DNA in the detection step c) can be performed using dUTP in place of one of the deoxyribonucleotide (dNTPs), where the product may then be optionally selectively degraded, inactivated or removed. This inactivation may be performed using an enzyme such as Uracil-DNA glycosylase, also known as UNG or UDG.

15 g) Repeating steps (a) to (e)  
Using the droplets containing enriched target DNA molecules obtained in (e) in a new step (a), the target DNA molecules may be further enriched.

20 h) Amplifying the enriched sample  
Using the enriched amplified target DNA molecules obtained in (e), the target DNA molecules may be further amplified using a general amplification such as MDA or a specific amplification such as PCR. For example, further amplification may be performed by collecting the water phase from the coalesced droplets obtained following general amplification in step e); adding reagents for general amplification to the water phase to form a reaction mixture; generating new droplets from the reaction mixture, and incubating the droplet emulsion at 30°C.

25 30 **Scheme for performing PINSSwift**

The scheme is outlined in Figure 4. (401): Determine the concentration of target DNA molecules in the original DNA sample. Dilute the sample until the expected average number of droplets containing a target molecule (positive droplets) is

less than 0.5. If only one target is enriched, the average number of positive droplets should be smaller resulting in a greater enrichment. Ideally, the entire sample of droplets should contain 2-100,000 positive droplets per specific target, preferably 5-10,000 and more preferably 20-100. When two different specific

5 targets such as two different genes are to be enriched from the same sample, the ideal number of positive droplets is two times higher. If more than one sequence variant of the target is present, each variant counts as a separate target DNA molecule. Also, if a target DNA molecule is long, such as longer than 20,000 base pairs, more than one properly spaced nucleotide sequence for

10 detection of specific target may be used. Each of these nucleotide sequences counts as separate DNA molecules. The correlation between number of targets, the average number of positive droplets and the resulting enrichment after PINSswift is shown in Figure 3 using an example of one round of PINSswift, 20,000 droplets and 10 positive droplets per target.

15 The enrichment may be higher or lower than the expected enrichment of figure 3 due to bias of this amplification step. In our experience the resulting enrichment may be at least two fold higher. Also, if a larger number of droplets such as one million droplets are used, this will result in a greater enrichment. To achieve a great enrichment a smaller droplet size is therefore typically preferable. A typical

20 size of the droplets may be less than 100  $\mu\text{m}$  in diameter, such as 1-100  $\mu\text{m}$ , 5-50  $\mu\text{m}$  or preferably 10-30 $\mu\text{m}$ . The droplets may also be of varying sizes within the same sample.

The diluted target DNA molecules are mixed with reagents suitable for detection of the target sequences in the sample. If detection is performed using PCR,

25 these reagents may include a polymerase, target specific primers, probes such as Taqman probes, molecular beacons, or scorpion probes, SybrGreen, nucleotides, and salts.

Droplets containing mixed DNA and reagents for detection are now generated (403)

30 If detection is performed using PCR; the droplet sample is transferred to a PCR apparatus and PCR amplification is performed and the droplets are then sorted according to the presence (positive) or absence (negative) of the target molecule. Where fluorescence labelled oligonucleotides are used, the presence of

the target molecule can be detected by the fluorescence of the droplets (404 and 405).

The selected and sorted droplets will contain a small amount of target DNA molecules such as 2-1000 molecules and amplification of said molecules is 5 therefore necessary before the DNA can be used for downstream applications e.g. DNA sequencing. Due to the small amount of DNA, it may not be possible to generally amplify the DNA using standard procedures. Methods to improve the amplification include repeated steps of amplification and re-amplification and combinations of amplification in droplets and standard amplification (in one 10 reaction, excluding droplet formation).

Determine the abundance of the target DNA molecule. If this is sufficient for analysis such as sequencing, the selected sample may be used directly; otherwise additional rounds of PINSswift can be applied.

15

When it is necessary to perform further rounds of enrichment, it may be preferred to degrade or remove the DNA generated in the detection step. It may be preferred to use dUTP in the PCR reagents. The MDA reaction is then performed using standard dNTPs. After detection and physical selection of 20 droplets, the DNA generated in PCR can be degraded and the treated sample can be used for an additional round starting with dilution, detection and physical selection.

25

When a sufficient enrichment is reached using PINSswift, the droplets are fused (also termed coalesced or demulsified). The enriched DNA may also be further purified and may be further amplified using general or specific amplification such as MDA and PCR respectively.

30

The PINSswift procedure employs general amplification to the target DNA molecule after it has been sorted. This has several advantages over a procedure where the DNA is generally amplified in droplets before detection and sorting including 1) the chemicals and enzymes needed for general amplification do not need to be present in the detection step, and 2) as the volume of the sorted droplets is small, the detection chemicals will only have minimal influence on the

general amplification step. When general amplification is performed using Phi29 polymerase in a process where the DNA is amplified in droplets before detection, the polymerase will degrade oligonucleotides to be used for detection thereby preventing correct identification of droplets containing target DNA molecule.

5 Also, when PCR is subsequently used for detection, the reaction conditions are not optimal for general amplification, and the general amplification can be inhibited. In a process employing general amplification in droplets before detection, the detection chemicals could alternatively be added to the droplets after general amplification thereby avoiding the problems of incompatibility

10 between the detection and general amplification reactions. However, this complicates the process by requiring a droplet fusion step. By performing general amplification after detection and sorting, PINSSwift achieves efficient detection while keeping a good performance of the general amplification and without adding the complexity of droplet fusion.

15

## **II: Multiplex PINSSwift**

PINSSwift can be adapted to perform multiplex PINSSwift. Multiplex PINSSwift employs additional features that are designed to detect a 2nd consecutive sequence of at least 10 (or 15, more preferably 40) nucleotides in the sample of

20 mixed DNA molecules analysed, by amplification of this 2nd consecutive sequence with sequence specific primers to generate a 2nd target DNA molecule. If several droplets show specific detection (e.g. by fluorescence signal) from both the 1st and 2nd consecutive sequence then they have a high probability of being located on the same MDA-amplified target DNA molecule.

25

In addition to providing information concerning co-localisation of targets, multiplex PINSSwift provides simultaneous purification of up to thousands of different target molecule, each comprising more than 5,000 base pairs. Separate detection molecules can be provided to separate droplets or can be added as a

30 mixture.

## **III Samples analysed by PINSSwift and Multiplex PINSSwift**

### III.i Sample of mixed DNA molecules

PINSSwift may be applied to a sample of mixed DNA molecules that may

35 comprise a target DNA molecule. A sample of mixed DNA molecules comprises a

population of DNA molecules (e.g. chromosomal DNA molecules or plasmid DNA molecules) where the individual DNA molecules within the population differ by at least one nucleotide within a known consecutive sequence of at least 10 (or 15 or more preferably 40) nucleic acid base pairs in their DNA, such that a target 5 molecule comprising the known consecutive sequence differs from, and can be distinguished from, non-target molecules in the sample. The sample of mixed DNA molecules may additionally comprise single stranded RNA or DNA polynucleotides. The population of DNA molecules in the sample of mixed DNA molecules comprises the target DNA molecule.

10 The target DNA molecule can be in linear or circular forms. Circular DNA can occur naturally or can be obtained by cloning DNA into plasmids, fosmids, cosmids, BAC clones, or generated by ligation or through Cre/LoxP mediated recombination.

15 A target DNA molecule comprises one or more known consecutive sequence of at least 10 (or 15 more preferably 40) unique nucleic acid base pairs (or nucleotides). A target DNA molecule can be selected from a sample of mixed DNA molecules, by selecting for a target DNA molecule comprising this 20 consecutive sequence of at least 10, 15 or 40 nucleic acid base pairs (or nucleotides). The target DNA molecule can also be selected from the sample of mixed DNA molecules, by selecting for a target DNA molecule comprising at least two consecutive sequences of at least 10 (or 15 more preferably 40) nucleic acid base pairs (or nucleotides), wherein the two consecutive sequences are 25 comprised within a DNA molecule of 50 to 100,000 nucleic acid base pairs, preferably 150 to 3,000 nucleic acid base pairs, more preferably 150 to 1500 nucleic acid base pairs.

30 Typically, the frequency of the target DNA molecule in the sample of mixed DNA molecules is less than  $10^{-2}$ , it may for example lie between  $10^{-3}$  and  $10^{-9}$  (calculated as base pairs of target sequence divided by base pairs of total DNA in the sample). For illustration purposes, the sample could be DNA from a human swab and the target DNA could be a region of 30,000 base pairs containing a specified gene. In this case, the frequency of the target DNA molecule is 30,000 divided by approximately 3 billion base pairs (the human genome) which equals

$10^{-5}$ .

The method of the invention is particularly suitable where the frequency of the target DNA molecule in the sample of mixed DNA molecules is less than  $10^{-2}$ .  
5 The method of the invention is also suitable where the frequency of the target DNA molecule in the sample of mixed DNA molecules is less than  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$ ,  $10^{-6}$ ,  $10^{-7}$ ,  $10^{-8}$ ,  $10^{-9}$ ,  $10^{-10}$ , or  $10^{-11}$ . In many instances, the sample of mixed DNA molecules will be derived from a cell population comprising genomic DNA, while in other instances the sample may be derived from samples where the  
10 DNA molecules are of diverse origin, such as environmental samples. Irrespective of its source, the frequency of the target DNA molecule is defined as the number of specific target DNA base pairs divided by the number of total base pairs in the sample. The frequency of the target DNA molecule in the sample of mixed DNA molecules is determined by making a dilution series in triplicate,  
15 detecting the presence or absence of target and determining the number of targets using, for instance, most probable number methods. The frequency of the target molecule can also be determined using qPCR or digital droplet PCR [3]. The concentration of DNA is measured and the number of total genome equivalents is determined by dividing this concentration by the average  
20 molecular weight of the genome.

### III.ii Source of the sample of mixed DNA molecules

According to one embodiment of the present invention, the target DNA molecule is derived from the genome of a cell, where the genome may be either  
25 chromosomal or extra-chromosomal DNA, or from DNA present outside of a cell such as from a virus or cell free DNA. Further, the target DNA molecule may be derived from a cell, where the cell is selected from amongst a microbial cell, a plant cell, an animal cell, a protist cell, a fungal cell, or a mammalian cell. The mammalian cell may be a human cell. The microbial cell may be a bacterial cell,  
30 a yeast cell or a fungal cell. Furthermore, the target DNA molecule may be derived from a fungal mycelium or fungal spores.

When the target DNA molecule is derived from one or more cell, the cell(s) may be part of a multicellular tissue or multicellular organism.

Furthermore, the target DNA molecule may be derived from one or more viral particles, where the virus has an RNA or DNA genome. Alternatively the target DNA molecule may be derived from a host genome comprising integrated DNA derived from a virus. The target DNA molecule may also be derived from a 5 bacteriophage.

Irrespective of the derivation of the target DNA or RNA molecule, the target DNA or RNA molecule is present in a sample of mixed DNA molecules, where the mixed DNA molecules may be derived from a sample collected from the 10 environment, for example a sample of soil, water or air. Alternatively, the sample may be derived from a multicellular organism, such as a mammal, for example an animal or a human subject. When the sample is derived from a mammal, the sample (for example a biopsy) may be derived from a body fluid (e.g. blood, plasma, serum, lymph and urine), from faeces or from a body tissue 15 or organ. The multicellular organism from which the sample is derived may be a living or may be a dead organism.

### III.iii Preparation of the sample of mixed DNA molecules

The sample of mixed DNA molecules comprising the target DNA molecule may be 20 prepared from a sample collected from nature or from an organism (e.g. a biopsy). Methods for selective extraction of DNA molecules are known in the art [4]. When the target DNA molecule is derived from a cell, the step of cell disruption or cell permeabilisation is normally required in order to release total nucleic acid molecules (including DNA or RNA) from a cell, this step preceding 25 the subsequent step of selective extraction of DNA molecules.

Where the target DNA molecule is derived from an RNA genome, the RNA genome or parts thereof are first reverse transcribed to provide a cDNA molecule, where the nucleotide sequence of the cDNA corresponds to (is a 30 reverse transcript of) the RNA genome.

Before generation of droplets, a compound such as a protein, or a compound comprising a DNA or RNA fragment, such as a DNA fragment coupled to a protein, may be bound to the target nucleotide in order to enable or facilitate 35 subsequent detection. Unbound compounds may be washed from the sample

before generation of droplets. Useful proteins include antibodies, single-chain variable fragment and enzymes. Useful compounds include DNA binding proteins such as regulatory proteins and histones.

5 The DNA sample may be circularized before the generation of droplets. Some general amplification methods, including the Phi polymerase catalyzed MDA reaction, are more efficient on circular templates. By circularizing the template, the DNA from the original nucleotide sample may therefore be favoured over non template reaction or templates stemming from the detection reaction and bias 10 against target can thereby be reduced.

The sample may be diluted before formation of droplets. If for example 1 million droplets are used for the enrichment of a gene from human chromosomal DNA 100 pg human DNA may be used to obtain 30 copies of target DNA molecule in 15 the 1 million droplets. As a copy of the human genome contains approximately 100,000 fragments of 30,000 base pairs, each droplet may therefore contain an average of 3 fragments. The approximately 30 droplets containing the target DNA molecule will therefore also contain non-target fragments. If the DNA is sequenced after the enrichment is finalized, the non-target fragments may also 20 be sequenced but, as the non-target fragments are likely to be different from each other, the sequence of the target DNA molecule is clearly distinguishable. For more complex samples such as a low copy virus in a human sample, the number of non-target fragments per target fragments in a positive droplet may be even higher such as higher than 100:1 or 10,000:1.

25 The total number of target DNA molecules is typically between 5 and 1,000, such as 5-100,000. When only one variant of a target DNA molecule is needed, the preferred number of total target DNA molecules added to the droplets is low such as 20-100 to achieve a high degree of enrichment. When several different 30 variants are needed, such as in the analysis of inhomogeneous tumour DNA, a higher number of total target DNA molecules may be preferred such as 100-10,000.

### III.iv Generation of droplets

Methods of the invention include forming multiple droplets where the droplets each contain less than 0.01 specific target molecules on average. In the preferred embodiment the distribution of specific target molecules follows Poisson distribution. Some droplets may contain non-target molecules present at 5 10 fold or higher concentrations as compared to the target molecule, while other droplets may contain only the target molecule.

Generally, droplets can be formed by a variety of techniques such as those described in [5-7]. Methods of the invention may involve forming a two-phase 10 system comprising aqueous droplets surrounded by an immiscible carrier fluid. In a preferred embodiment, the aqueous sample within the droplet is prepared by preparing a mixture suitable for specific detection of target DNA molecules containing sample DNA and e.g. a polymerase, primers, probes and buffer solutions.

15 The resulting mixture is used as the aqueous sample in two-phase liquid droplet formation using two immiscible liquid phases. Aqueous droplets are either generated in an apparatus having means for creating a vortex/turbulence in a sample comprising two immiscible liquid phases in controlled environments, creating droplets of liquid (phase 1) in a 2nd phase liquid by controlling the 20 mechanical parameters and thereby also the liquid volume of each generated droplet or by a means for extruding droplets of one liquid phase in a 2nd immiscible liquid phase, where the so formed droplets remain discrete and wherein the volume of the droplet is controlled by the diameter of the means for droplet extrusion.

25 The carrier fluid is one that is immiscible with the sample fluid. The carrier fluid can be a non-polar solvent, decane, fluorocarbon oil, silicone oil or any other oil (for example mineral oil).  
In certain embodiments, the carrier fluid contains one or more additives such as agents which increase, reduce, or otherwise create non-Newtonian surface 30 tensions (surfactants) and/or stabilize droplets against spontaneous coalescence or contact.

#### **IV Methods of detecting the target sequence**

Methods of the invention further involve detection of the target nucleic acid 35 molecule within the droplets containing target DNA molecules. In certain

embodiments the detection involves amplification of a part of the target molecule. The amplification reaction, that is suitable for amplifying nucleic acid molecules, includes polymerase chain reaction, or nested polymerase chain reaction including or excluding probes such as Taqman probes, Scorpion probes,  
5 Molecular Beacon probes, Padlock probes, Molecular inversion probes and any other probe that functions by sequence specific recognition of target DNA molecule by hybridization and result in increased fluorescence on amplification of the target sequence.

Methods according to the invention also include methods wherein detection is  
10 based on fluorescence from optically labelled probes such as fluorescently labelled probes wherein the target is not amplified. In this case, the DNA is denatured e.g. by increasing temperature to around 95°C and the probe is subsequently allowed to anneal to the target, resulting in activation of the probe or probes. Such optically labelled probes can be Molecular Beacons, where a  
15 single-stranded bi-labelled fluorescent probe is held in a hairpin-loop conformation of around 20 to 25 nt by a complementary stem sequences of around 4 to 6 nt. Due to the loop-structure the desired fluorochrome attached to one end of the sequence is in close proximity of the light quencher attached to the other end. When the structure is released during denaturing and then re-annealed, the probe anneals to an amplified target. When annealed, the hairpin structure is no longer maintained, and the quencher no longer quenches emitted light from the fluorochrome. The optically labelled probes can also be FRET (fluorescence resonance energy transfer) probes.  
20  
25

The specific target sequence may also be detected using nucleic acid stains such as cyanine dyes.

#### **V Methods of physically selecting droplets based on presence of the target sequence**

To selectively separate droplets comprising a detectable target DNA molecule  
30 from droplets wherein the target is not detected, a variety of different methods for physical selection of droplets or droplet sorting can be employed including steering, heating, electro-wetting, magnetic activated sorting, acoustic waves [6], and dielectrophoretic sorting [8]. Such physical selection can be carried out using an apparatus with means for receiving droplets of an aqueous liquid that  
35 are suspended in a 2nd immiscible liquid (e.g. droplets from specific detection

step), and means for passing each droplet past a detection unit capable of detecting a component in said droplet, and means for addressing said droplet for delivery to a selected compartment as determined by the presence or absence of the detectable component and means for delivering said droplet to the selected 5 compartment.

#### **VI Methods of General Amplification of DNA suitable for PINSSWIFT**

A range of different approaches have been suggested for general amplification of DNA, such as randomly degenerate primed PCR, linker ligation PCR, or, 10 Degenerate Oligonucleotide Primed (DOP) PCR and Multiple Displacement Amplification (MDA). MDA is reported to be efficient in performing whole-genome amplification (WGA) of even small amounts of DNA (but with a lower limit of circa 1ng). Compared with more traditional PCR-based WGA methods, MDA generates DNA molecules with a higher molecular weight, having better genome 15 coverage. MDA employs a strand displacement polymerase that possesses two enzymatic activities: DNA synthesis (polymerase) and an exonucleolytic activity that degrades single stranded DNA in the 3'- to 5'-direction, as exemplified by bacteriophage phi29 polymerase, that belongs to eukaryotic B-type DNA polymerases (UniProtKB/TrEMBL: Q38545). Other useful polymerases include 20 BstI polymerase. The amplification may be performed in droplets to increase the local concentration of DNA. In some embodiments, the amplification is performed at around 30°C for four hours, preferably less than one hour.

As the amount of target DNA molecule from the detection and sorting step may 25 be as low as five or fewer molecules, typical general amplification reactions will either produce no product or produce background reaction (no template reaction).

It was found that background reaction is prevented and sensitivity is increased when amplification is carried out in droplets such as water-in-oil droplets of 1 pL 30 to 1 nL thereby making PINSSwift possible at low numbers of DNA target molecules such as less than 100 molecules corresponding to approximately 3 fg of DNA. The reaction is improved even when the droplets are of variable sizes such as droplets produced by vortexing. The droplets are subsequently coalesced (also termed demulsified) either in one step or gradually to form a solution 35 containing the enriched target DNA molecules.

## **VII Methods of removing the detection signal molecules after physically selecting the droplets**

After physically selecting the droplets based on the presence of the specific target sequence, it may in some cases be necessary to remove a detection signal, such as a PCR product. Several methods for removing such signals are known in the art. If dUTP has been used in the detection reaction, the detection molecule may be removed using uracil-DNA N-glycosylase [9]. Alternatively, as the molecules produced by general amplification are significantly longer than the detection molecules, the detection molecules can be separated using methods based on size separation such as size exclusion based on differential binding affinity of small and large DNA to silica particles [10]. Such silica surfaces have limited binding efficiency to DNA fragments smaller than 100 bp, and consequently only DNA fragments smaller than 100 bp will efficiently be discarded, when silica based purification is applied. In some applications, however, it may not be necessary to remove the detection molecule. For instance, since Phi polymerase and some other polymerases have low activity on DNA molecules shorter than 1000 base pairs, it may not be necessary to remove the detection molecule if the step following PINSSwift enrichment is a general amplification since the detection molecule will only be amplified to limited extent, if at all amplified, compared to the actual targeted larger DNA molecule.

## **VIII Sequence determination of the target DNA molecule**

When a target DNA molecule is sufficiently enriched and generally amplified, the nucleotide sequence of the molecule including the target nucleotide sequence can be determined. The enriched target DNA molecule can be sequenced using e.g. Sanger sequencing, Emulsion PCR, Shotgun sequencing, SOLiD sequencing, bridge PCR, Ion Torrent sequencing, Polony sequencing, Pyrosequencing, Sequencing by synthesis, DNA nanoball sequencing, Heliscope single molecule sequencing, Nanopore DNA sequencing, Tunnelling currents DNA sequencing, Sequencing by hybridization, Sequencing with mass spectrometry, Transmission electron microscopy DNA sequencing, RNAP (RNA Polymerase) sequencing or Single-molecule real-time sequencing.

35 **IX Applications of PINSSwift**

### IX.i Research and development applications of PINSswift

Use of PINSswift to isolate or enrich target DNA molecule in mixed sample of DNA molecules extracted from samples collected from the environment, or from clinical samples provides direct access to the genome, or parts thereof, that 5 cannot be analysed by other methods because of the sample complexity. PINSswift is particularly useful for isolating or enriching DNA molecules involved in hereditary diseases, cancer, and infectious diseases.

Multiplex PINSswift is particularly useful for simultaneous isolation or enrichment of more than one target DNA molecule from samples comprising several target 10 DNA molecules, such as a sample to be analysed for the DNA sequence of more than one virus, more than one hereditary disease or more than one cancer related gene.

PINSswift is also particularly useful for obtaining target DNA molecule sequence information from samples where only a small part of the sequence is known prior 15 to enrichment, since the technique only requires a small part of the target DNA molecule sequence to be known in order to perform the detection step and furthermore takes advantage of high fidelity of polymerases (like Phi polymerase) in the MDA amplification to generate large amplified DNA molecules of up to 100,000 bp.

### IX.ii Sample preparation for sequencing

When performing sequencing of large DNA molecules such as genomes, the sequence information obtained will often contain gaps where the sequenced molecules do not overlap or cannot be assembled. PINSswift is particularly useful for gap closing of DNA sequences, as PINSswift will retrieve up to 100,000 bp of 25 sequence surrounding the small detection area and can therefore be designed to cover the unknown gap region.

PINSswift is also particularly useful when sequencing samples containing more than one variant of a target DNA molecule sequence such as chromosomal DNA containing two alleles of a gene. In this case, droplets containing each detection 30 sequence are collected in separate compartments to ensure that only one copy of the sequence is present. The droplet may subsequently be barcoded

separately to enable the separate sequencing of each variant of the target DNA molecule sequence.

### IX.iii Diagnostic applications of PINSSwift

5 PINSSwift may be used to analyse target DNA molecule in samples derived from multicellular organisms, such as a biopsy or a sample of body fluid or faeces obtained from a subject (e.g. human or animal subject), for the diagnosis or monitoring the progress of a medical indication or disease.

10 Diagnosis of a wide range of medical indications in a subject such as a disease caused by an infectious agent (e.g. micro-organism or virus) can be assisted by the isolation or enrichment and detection of a target DNA or RNA molecule that is derived from the genome of the infectious agent by PINSSwift or Multiplex PINSSwift, where the target DNA molecule is detected in a sample of mixed DNA molecules derived from a biopsy or sample of body fluid obtained from a patient.

15 Use of PINSSwift in target DNA molecule isolation provides the additional feature, that additional diagnostic features of the disease can be determined. For example, where the genome of the infectious agent comprises resistance genes that confer resistance to certain therapeutic agents, enrichment of a resistance gene may also retrieve the region surrounding the resistance gene and may 20 provide information about the specific infectious agent carrying the resistance.

25 PINSSwift may be applied to diagnostics such as detection of presence of an infectious agent such as a prokaryotic organism and antibiotic resistance. Such cases can be the presence of methicillin resistance (MR) in *Staphylococcus aureus* (SA). The combination of (MRSA) is a well-known problem in hospitals and similar facilities, whereas MR may not cause comparable problems if it is not present in SA. PINSSwift can be used for detection of co-existence of MR and SA if a method such as duplex detection is applied. Such duplex detection can be a dual-reporter detection system, where one detection system is used to monitor 30 the presence of one event such as MR whereas another detection system is used to monitor the presence of another event, such as presence of SA within the same droplet. When both MR and SA are present in the same droplet, the two loci are likely to be localized on the same DNA fragment. Moreover, PINSSwift

can be applied to selectively retrieve additional DNA sequences from the host genome based on the presence of genes such as MR or SA. Thus, it can be used to provide additional sequence information from the genome of the infectious organism, which is of importance for bacterial typing.

5

PINSswift may also be used to assist diagnosis of a disease caused by, or originating from the presence of a viral agent in a subject. Using multiplex PINSswift, the presence or absence of viral DNA at a known integration site can be determined by tracking the co-enrichment of the viral DNA and integration site DNA sequence.

10

PINSswift may be applied to multiple genes within a single genome or within multiple genomes, which may be detected by any suitable molecular detection method. If required, a series of multiplex PCR reactions can be carried out and 15 differentiated using specific dyes for each reaction. This can be done by introduction of detection systems such as probes such as Taqman probes, Scorpion probes, Molecular Beacon probes or similar. Moreover, PINSswift can also be applied to a series of genes which are not necessarily differentiated at the point of detection. Differentiation can be applied after sequence retrieval, 20 using such methods as bar-coded PCR or similar.

### **X PINSswift and bias in amplification**

25

PINSswift is based on specifically selecting samples where amplification of a desired DNA region from a complex, mixed DNA sample has occurred. Although 25 Phi29 polymerase based amplification (MDA) has been described repeatedly as the most reliable genome amplification currently available, it is known to introduce significant bias. Pan et al. [18] states in general terms that a highly specific whole genome amplification (WGA) of complex DNA pools which avoids amplification bias remains a challenge. Moreover, similar observations are seen 30 with alternative amplification methods such as DOP-PCR and random priming PCR. These two amplification methods are described as being much less efficient at reproducing the locus representation [19], resulting in even more biased amplification products. While bias is seemingly unavoidable, regardless of the amount of reaction template [20] present, the amount of template independent 35 product (TIP) or bias introduced during amplification is seemingly correlated

negatively to the amount of DNA template in the reaction and has in some studies been documented to represent 70–75% of the total yield [18]. Whole genome amplification is applied to amplify DNA in the PINSswift process and it would therefore be expected that these general challenges relating to bias in 5 genome amplifications would also apply to PINSswift. As a procedure including WGA it would thus be expected that significant bias against the target DNA molecule should be observed. Hence a procedure such as PINSswift employing WGA would not have been considered as a method capable of enriching for a specific region of DNA in a mixed sample.

10 Surprisingly, the challenge of TIP/bias is not seen when applying the PINSswift technique, because the step of general amplification (e.g. MDA) is performed in at least  $1.2 \times 10^6$  droplets per 5 $\mu$ L MDA reaction mixture, each droplet having a volume of 5pL or less; where the total amount of DNA molecule amplified is less than 300 fg (more commonly  $\leq$ 50fg). Under these conditions, amplification of 15 the DNA molecules (normally 1 target DNA molecule and  $\leq$ 2 non-target molecules in each droplet) is reagent-limited due to the pL volume of MDA reagents in each droplet. As a result, the target DNA molecule is effectively amplified while non-target DNA molecules, that might otherwise be preferentially amplified by Phi29 polymerase, are not overrepresented in the final product. As 20 stated above, performing general amplification in these droplets also serves to eliminate amplification of contaminating DNA molecules (see Example 5). The successful implementation of the PINSswift method was not expected because the amplification of the target DNA molecules in the general amplification step requires a combination of pL reaction volumes and a template of as low as one 25 to two DNA molecules. Such conditions for general amplification have never previously been attempted, not least because ng amounts of template are recommended for Phi polymerase amplification, and because effective enzyme concentrations are known to fall dramatically in small volume reactions due to surface adsorption at the oil-water interface (see Example 6).

30 The minimal negative TIP/bias observed in the PINSswift system is significantly lower than the overall gain obtained from the amplification process and the net result is therefore a substantial enrichment of the target DNA molecule.

35 **EXAMPLES**

**Example 1. Enrichment of two regions of DNA from *Escherichia coli* mixed into human genomic DNA and subsequent sequencing**

The procedure employed is shown schematically in Figure 5.

DNA sample and detection reagents

5 1  $\mu$ L of a DNA sample containing 0.2 pg/ $\mu$ L of chromosomal *Escherichia coli* DNA and 1 pg/ $\mu$ L of HeLa human chromosomal DNA was mixed with standard PCR reagents to a total volume of 10  $\mu$ L. One region (E. coli – amplicon1) was targeted by primers "MB8 Fw1 tm60" and "MB8 Rev1 tm60" annealing to *Escherichia coli* AspartokinaseI (*thrA*) together with molecular beacon probe "MB 8.7" annealing to the region between the two PCR primers. The second region (E. coli – amplicon2) was targeted by primers "ThrB-fw" and "ThrB-Re" annealing to *Escherichia coli* Homoserine kinase (*thrB*) together with Beacon-ThrB annealing to the DNA between the two primers.

Preparation of droplets

15 5  $\mu$ L PCR reaction + 50  $\mu$ L droplet oil (4% v/v ABIL EM90, 0.05% v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology; Sigma-Aldrich; # M5904]) was mixed on a vortex mixer generating approximately 2 million water-in-oil droplets of different sizes with an average volume of approximately 2.5 pL. Assuming an average fragment length of 50 kilobases, each droplet contains on 20 average  $1.8 \times 10^{-3}$  *Escherichia coli* fragments hereof approximately  $1.7 \times 10^{-5}$  target fragments and  $9.1 \times 10^{-3}$  human non-target fragments.

PCR

Emulsion PCR was performed with the following cycling conditions: 95°C 2 min, 3x(95°C 3 seconds, 56°C 15 second), 25°C 30 seconds.

25 Detection and Selection

The sample was transferred to a microfluidic chip with a sorting channel diameter of 100  $\mu$ m using a syringe pump and transported into a t-junction sorting zone on the chip. The droplets were visualized using a camera and bright fluorescent droplets (positive droplets) were physically sorted into one channel 30 while non-fluorescent or weakly fluorescent droplets were discarded and left unsorted and were subsequently discarded. When all droplets were sorted to

either gate, the positive droplets (a total of 31) followed by an oil phase were removed from the chip using vacuum from a syringe pump followed by gravity flow and transferred to a 200 $\mu$ L tube. The total amount of the DNA molecules in the detected droplets, following coalescence, is estimated to be 1.7 fg, assuming  
5 that the average size of a detected target DNA molecule is 50,000 base pairs.

#### General amplification

10  $\mu$ L of 10 mM Tris-HCl, pH 8 was added to the oil containing the sorted positive droplets. The droplets were coalesced by adding 40  $\mu$ L perfluorooctan-1-ol, vortexing and centrifugation. The 10  $\mu$ L water phase was transferred to a new  
10 tube containing reagents for MDA amplification, including Phi29 polymerase, dNTPs, random hexameric primers and suitable salts to reach a total volume of 20  $\mu$ L. 50 $\mu$ L oil (4% v/v ABIL EM90, 0.05% v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology; Sigma-Aldrich; # M5904]) was added and the two phases were mixed on a vortex mixer generating approximately 2 million  
15 water-in-oil droplets of different sizes with an average volume of approximately 5 pL. After two hours, the emulsion was broken by adding 40  $\mu$ L perfluorooctan-1-ol, vortexing and centrifugation, the water phase was transferred to a new tube and the Phi polymerase amplification process was repeated, and the emulsion was coalesced. The final water phase contained 370 ng/ $\mu$ L total DNA in  
20 20  $\mu$ L. The number of target DNA molecules was quantified using qPCR with primers annealing to a region adjacent to the primer pair used for detection (primers MB10.1fw and MB10.1rev).

#### Results

25 After enrichment as described above, the final sample contained  $1.1 \times 10^{10}$  target copies per  $\mu$ L as quantified using qPCR as described above or 30000 targets per pg. This corresponds to an enrichment of approximately 1090 fold (calculated on a targets / pg total DNA basis). The sample was sequenced using Sanger sequencing and next generation sequencing (Illumina 150 bp paired-end sequencing) and both the expected *Escherichia coli* sequence were confirmed.

30 **Example 2 – Identifying a point mutation in mixed human DNA.**

The procedure employed is shown schematically in Figure 5.

DNA sample and detection reagents

Purified human DNA containing a point mutation at codon 12 (sequence GAT) in gene Kras (JX512447) was mixed 1:100 with a corresponding sample without the mutation in codon 12 (sequence GGT) (mutation frequency 1%). 1  $\mu$ L of a 5 DNA sample containing 330 ng of the mixed chromosomal DNA was mixed with standard PCR reagents to a total volume of 20  $\mu$ L. Primers Kras-cdn12-fw (TAGTGTATTAACCTTATGTG - SEQ ID No: 19) and Kras-cdn12-re (TTACCTCTATTGTTGGAT - SEQ ID No: 20) annealing to Kras were used along with Taqman probe taqMan-Kras-cdn12 annealing to the region between the two 10 PCR primers at the site of the mutation.

Preparation of droplets

20  $\mu$ L PCR reaction + 60  $\mu$ L droplet oil (4% v/v ABIL EM90, 0.05% v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology; Sigma-Aldrich; # M5904]) was mixed on a vortex mixer generating approximately 8 million water-in-oil droplets of different sizes with an average volume of approximately 2.5 pL. 15 Assuming an average fragment length of 50 kilobases, each droplet contains on average 760 fragments hereof approximately  $1.2 \times 10^{-4}$  target fragments corresponding to 1000 positive droplets in total.

PCR

20 Emulsion PCR was performed with the following cycling conditions: 95°C 2 min, 3x(95°C 5 seconds, 56°C 15 second, 72°C 15 seconds), 25°C 30 seconds.

Detection and Selection

The sample was transferred to a microfluidic chip with a sorting channel diameter of 105  $\mu$ m using a syringe pump and transported into an x-junction 25 sorting zone on the chip. The droplets were visualized using a camera and bright fluorescent droplets (positive droplets) were physically sorted into one channel while non-fluorescent or weakly fluorescent droplets were sorted into the other channel. When all droplets were sorted to either gate, the positive droplets (a total of 950) followed by an oil phase were removed from the chip using vacuum 30 from a syringe pump and transferred to a 200 $\mu$ L tube. The total amount of the DNA molecules in the selected positive droplets is estimated to be 50 fg,

assuming that the average size of a detected target DNA molecule is 50,000 base pairs.

General amplification

10  $\mu$ L of 10 mM Tris-HCl, pH 8 was added to the oil containing the sorted 5 positive droplets. The droplets were broken by adding 200  $\mu$ L mineral oil, vortexing and centrifugation. The 10  $\mu$ L water phase was transferred to a new tube containing reagents for MDA amplification, including Phi29 polymerase, dNTPs, random hexameric primers, dithiothreitol (DTT) and suitable salts to reach a total volume of 20  $\mu$ L. 60  $\mu$ L droplet oil (4% v/v ABIL EM90, 0.05% v/v 10 Triton X100, 96% v/v Mineral oil [light oil for molecular biology from [www.sigmaaldrich.com](http://www.sigmaaldrich.com), product M5904]) was added and the two phases were mixed on a vortex mixer generating approximately 4 million water-in-oil droplets of different sizes with an average volume of approximately 5 pL. After two hours at 30°C, the emulsion was broken by adding 120  $\mu$ L mineral oil, vortexing and 15 centrifugation, the water phase was transferred to a new tube and the emulsion was broken. The final water phase contained 0.098 ng/ $\mu$ L total DNA in 20  $\mu$ L.

10  $\mu$ L of the above enriched DNA sample was amplified using Phi29 polymerase in a total of 20  $\mu$ L as described above. The resulting enriched sample contained 350 ng/ $\mu$ L total DNA (sample A).

20 10 pg of the first enriched sample (before the second round of Phi29 polymerase amplification) was subjected to a second round of enrichment by performing emulsion PCR, selecting and sorting positive droplets. Approximately 400 droplets were selected and sorted as described above and the sorted droplets were further amplified resulting in a sample of 412 ng/ $\mu$ L total DNA in 20  $\mu$ L (B- 25 sample).

The number of target DNA molecules in samples A and B were quantified using qPCR with primers annealing to a region adjacent to the primer pair used for detection.

Results

30 After enrichment as described above, the final Sample A contained  $1.7 \times 10^7$  target copies per  $\mu$ L as quantified using qPCR as described above or 40 targets

per pg. This corresponds to an enrichment of 13000 fold (calculated on a targets / pg total DNA basis).

5 The final Sample B contained  $1.2 \times 10^{10}$  target copies per  $\mu\text{L}$  as quantified using qPCR as described above or 28000 targets per pg. This corresponds to an enrichment of 70 fold from Sample A and 910,000 fold from the initial sample (calculated on a targets / pg total DNA basis).

The sample was sequenced using Sanger sequencing and next generation sequencing (Illumina 150 bp paired-end sequencing) and the expected sequence including the point mutation was confirmed.

10 **Example 3. Identification of the sequence and integration point of a virus integrated into human DNA from 50 pg of total DNA**

The procedure employed is shown schematically in Figure 7.

DNA sample and detection reagents

15 1  $\mu\text{L}$  of a DNA sample containing 50 pg/ $\mu\text{L}$  of HeLa human chromosomal DNA containing integrated HPV18 virus in 20  $\mu\text{L}$  was mixed with standard PCR reagents to a total volume of 10  $\mu\text{L}$ . Primers "HPV18 104 fw1" and "HPV18 105 rev1" annealing to HPV18 were used along with molecular beacon probe "HPV18 106p" annealing to the region between the two PCR primers. dUTP was added instead of dNTP to enable the subsequent removal of PCR product from the 20 reaction.

Preparation of droplets

25 10  $\mu\text{L}$  PCR reaction + 30  $\mu\text{L}$  mineral oil droplet oil (4% v/v ABIL EM90, 0.05% v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology from www.sigmaaldrich.com, product M5904]) was mixed in a microfluidic chip for droplet generation generating approximately 200,000 water-in-oil mono disperse droplets of with an average volume of approximately 50 pL. Assuming an average fragment length of 50 kilobases and four copies of HPV18 per genome, each droplet contains on average 4.6 DNA fragments hereof approximately  $3.0 \times 10^{-4}$  target fragments.

30 PCR

Emulsion PCR was performed with the following cycling conditions: 95°C 2 min, 3x(95°C 5 seconds, 56°C 15 second, 72°C 15 seconds), 25°C 30 seconds.

Detection and Selection

The sample was transferred to a microfluidic chip with a channel diameter of 100 µm using a syringe pump and transported into a t-junction sorting zone on the chip. The droplets were visualized using a camera and bright fluorescent droplets (positive droplets) were physically sorted into one channel while non-fluorescent or weakly fluorescent droplets were sorted into the other channel. When all droplets were sorted to either gate, the positive droplets (a total of 57) followed by an oil phase were removed from the chip using vacuum from a syringe pump and transferred to a 200µL tube. The total amount of the DNA molecules in the selected positive droplets is estimated to be 14.2 fg, assuming that the average size of a detected target DNA molecule is 50,000 base pairs.

15 General amplification

10 µL reagents including UDG (uracil-DNA N-glycosylase) and buffers for removal of dUTP containing PCR products was added to the oil containing the sorted positive droplets. The droplets were coalesced by adding 200 µL mineral oil, vortexing and centrifugation. The 10 µL water phase was transferred to a new tube and incubated for 10 minutes at 37°C followed by 1 minute at 95°C for inactivation of UDG. Reagents for MDA amplification, including Phi29 polymerase, dNTPs, random hexameric primers and suitable salts were added to reach a total volume of 20 µL. 60 µL mineral oil was added and the two phases were mixed on a vortex mixer generating approximately 4 million water-in-oil droplets of different sizes with an average volume of approximately 5 pL. During incubation at 30°C, the droplets gradually fused resulting in an increased droplet size after the total of 4 hours incubation. The final water phase contained 275 ng/µL total DNA in 20 µL. The number of target DNA molecules was quantified using qPCR with primers annealing to a region adjacent to the primer pair used for detection (primers: HPV5901f (GTTTAGTGTGGGCCTGTGC) and HPV5994r (GGCATGGGAACTTCACTGT).

Results

After enrichment as described above, the final sample contained  $1.5 \times 10^9$  target copies per  $\mu\text{L}$  as quantified using qPCR as described above or 5500 targets per pg. This corresponds to an enrichment of approximately 1100 fold (calculated on a targets / pg total DNA basis). The sample was sequenced using Sanger sequencing and next generation sequencing (Illumina 150 bp paired-end sequencing) and the sequences of the regions of four integrated copies of HPV18 including integration sites were determined.

**Example 4. General amplification of a DNA target in droplets eliminates amplification of contaminating DNA**

10 All Phi29 polymerase amplification reaction mixtures were prepared in 25  $\mu\text{L}$  volume using the RepliG single cell kit from Qiagen according to the manufacturer's recommendations. DNA purified from *E. coli* was used as template for the reactions. Duplicate samples were prepared, where the amplification reaction in one sample was allowed to proceed in the 25  $\mu\text{L}$  volume  
15 (standard conditions); while the duplicate samples was first converted to droplets and then the reaction was allowed to proceed. When the reaction was performed in droplets, 50  $\mu\text{L}$  of mineral oil droplet oil (4% v/v ABIL EM90, 0.05% v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology from [www.sigmaaldrich.com](http://www.sigmaaldrich.com), product M5904]) was added and vortexed for 1 minute  
20 generating approximately 5 million water-in-oil droplets of different sizes with an average volume of approximately 5 pL. The reactions were incubated for 16 hours at 30°C. The emulsion was then coalesced by adding 50  $\mu\text{L}$  perfluorooctan-1-ol, vortexing and centrifugation and the 20  $\mu\text{L}$  water phase was transferred to a new tube. 2  $\mu\text{L}$  of each reaction was loaded onto a 0.7% agarose gel. The following amplification reactions were performed: 1) 1 ng *E. coli*  
25 DNA template, no droplets, 2) No DNA template, no droplets, 3) 1 ng *E. coli* DNA template, reaction in droplets, 4) No DNA template, reaction in droplets, 5) 1 ng *E. coli* DNA template, no droplets, no Phi29 polymerase.

Results

30 Figure 8 shows an agarose gel of the size separated amplification products of the Phi29 polymerase amplifications. Significant amounts of amplified DNA can be seen in lanes 1-3 corresponding to an efficient amplification. As the reaction in lane 3 does not contain template DNA, the amplified DNA is likely to come from

contaminating DNA in the reaction. When the non-template reaction was performed in droplets, no amplified DNA is visible on the gel (lane 4).

To investigate whether the amplified DNA from the Phi29 polymerase amplification reactions contains target DNA molecules (E. coli), quantitative PCR 5 was performed on the amplified samples using E. coli specific primers C-Thr 1325f and C-Thr 1485r and the following cycle conditions: 94°C, 15 sec/60.5°C, 15 sec/72°C, 15 sec.

As can be seen from Figure 9, only the Phi29 polymerase reactions comprising E. coli DNA template (1 and 3) produced specific product using the E. coli primer 10 set. This was the case for both the droplet and the non-droplet reaction. The Phi polymerase reaction without template and without droplets (Figure 8, Lane 2) did not result in any specific qPCR amplification although large amounts of DNA were present (Figure 9, 2) suggesting that the DNA generated by the Phi29 polymerase reaction was amplified from contaminating non-E. coli DNA.

15 The analysis further demonstrates that on average 0.2 fg template per droplet is sufficient for Phi29 polymerase mediated general amplification, when the reaction is performed in droplets having an average volume of 5 pL, as is the case for sample 3 comprising E.coli template.

#### **Example 5. Enrichment and sequencing of the human *dusp3* gene**

20 The following experiment was performed to demonstrate the detection, enrichment and sequencing of a target DNA molecule in a sample of mixed DNA molecules, where the frequency of the target DNA molecule in the sample of mixed DNA molecules was between  $10^{-5}$  and  $10^{-7}$  (calculated as base pairs of target sequence divided by base pairs of total DNA in the sample) and the target 25 DNA molecule comprised from between 10,000 to 100,000 nucleic acid base pairs.

The experiment was performed according to the method of the invention, where the sample of mixed DNA molecules was detected by PCR in droplets, where the frequency of target DNA molecules was between  $10^{-4}$  to  $10^{-5}$  target DNA 30 molecules per droplet. Approximately 100 fg DNA was recovered from positive droplets comprising the target DNA molecule, and a sample thereof was

subsequently generally amplified. Amplification reaction was performed in droplets having a size of about 5pL.

DNA sample and detection reagents

1  $\mu$ L of a DNA sample containing 1 ng/ $\mu$ L of human chromosomal DNA (Jurkat) 5 was mixed with standard PCR reagents to a total volume of 20  $\mu$ L. Primers "Dusp3 fw3" [5'-AGATGGTTTGCCCGCTT - SEQ ID No. 21] and "Dusp3 rev3" [5'-TGCCACTTAGCAGAAGCAAC - SEQ ID No. 22] annealing to the human DUSP3 gene (GenBank accession number NM\_004090.3) were used along with Taqman probe "Dusp3 tp3-2 FAM" [FAM-CCACCTCATATGTGTGCTGCC-BHQ1 - SEQ ID No. 23] annealing to the region between the two PCR primers. dUTP was added instead of dNTP to enable the subsequent removal of PCR product from the 10 reaction.

Preparation of droplets

20  $\mu$ L PCR reaction and 30  $\mu$ L mineral oil droplet oil (4% v/v ABIL EM90, 0.05% 15 v/v Triton X100, 96% v/v Mineral oil [light oil for molecular biology; Sigma-Aldrich; # M5904]) were mixed in a microfluidic chip for droplet generation and approximately 6,500,000 water-in-oil mono disperse droplets were generated, where the average droplet volume was approximately 3 pL. Assuming an average fragment length of 50 kilobases and one copy of the DUSP3 gene per 20 genome, each droplet contains on average 2.6 DNA fragments hereof approximately  $4.2 \times 10^{-5}$  target fragments.

PCR

Emulsion PCR was performed with the following cycling conditions: 94°C 2 min, 40x (94°C for 3 seconds, 60°C for 30 seconds), 25°C for 30 seconds.

25 Detection and Selection

The sample was transferred to a fluorescence activated cell sorter (Biorad S3e). Positive droplets were physically sorted into one gate while non-fluorescent or weakly fluorescent droplets were sorted into the waste collection. When all droplets were sorted to either gate, the positive droplets (a total of 592) were 30 transferred to a 200  $\mu$ L tube.

General amplification

The sorted positive droplets were coalesced by removing excessive oil, adding 2 x 10µL 1H, 1H, 2H, 2H, Perfluorooctanol, and mixing and centrifuging the mixture. A 3 µL sample (approximate 30fg) of the water phase of the coalesced droplets was added to a 17 µL mix containing N6 hexamer primer, water, dNTP, 2.5 unit Phi29 polymerase. 20 µL 2% PicoSurf oil was added and the reactions were then vortexed for 2 minutes at 2600 rpm to generate water-in-oil droplets where the average droplet volume was approximately 5pL. The droplets were incubated at 30°C for 6 hours and then the enriched and amplified DNA in the coalesced droplets was analysed using qPCR and Tapestation (Agilent) measurements. According to these measurements, the 3 µL DNA sample, following enrichment and amplification, now contained 60,000 target molecules. Since the total DNA concentration was below 0.5 ng/µL, which is the lower limit of detection of the Tapestation, a second round of general amplification in droplets was therefore performed as above to increase the amount of total DNA.

Results

The result of the second amplification in droplets was as follows in Table 1:

Table 1

Sample	Cq	copies/ul	ng/ul	copies/ng	Enrichment*
2nd Phi 592 3ul	16,16	23187840	31	747995	2693

20 \*Enrichment is relative to the original Jurkat DNA sample.

After enrichment as described above, the final sample contained  $2.3 \times 10^7$  target copies per µL as quantified using qPCR. This corresponds to an enrichment of approximately 2700 fold (calculated on a targets / pg total DNA basis). The sample was sequenced using next generation sequencing (Illumina 150 bp paired-end sequencing). A 14.5 kb sequence contig [SEQ ID No. 24] covering the DUSP3 gene target region was generated from the NGS data.

**Example 6. Enzyme inactivation as a result of adsorption to the water/oil interface**

30 Introduction

This example shows the effect on droplet-based Phi29 polymerase amplification as a function of reduction in droplet size. It is expected that the Phi29 polymerase reaction will be inhibited in small droplets based on prior art where enzymes are described as being highly susceptible to adsorption to the 5 interphase between water and droplet oil [11]. Moreover, it has been thoroughly described, that the increased surface to volume ratio causes an increased inhibition when performing enzymatic reactions in droplets [12].

Due to progressive inactivation of enzymes resulting from nonspecific adsorption at the oil-water interface, Phi29 polymerase amplification in small size droplets is 10 expected to be inhibited. Nonetheless, droplets created in this experiment (in the femto-to-picliter range and, thus, expected to result in inhibition) seemingly do not inhibit the Phi29 polymerase amplification, as DNA yield from different sizes of Phi29 polymerase droplets are close to identical (Table 3 Results from Droplet Phi).

15 Procedure

Two parallel 20 $\mu$ l samples were prepared for Phi29 amplification as described in Example 5 where 3ng Jurkat DNA template (Thermofisher Scientific) was added to each of the reactions. 40  $\mu$ l 2% PicoSurf (Dolomite Microfluidics) was then added to the samples and both were vortexed at different velocities as described 20 in the table below. "Hand-pulled on Eppendorf holder" is a method where a 1.5mL Eppendorf-tube containing the reaction mixture is held at an angle of approx. 45 degrees and forced over the top of a 16-hole 1.5mL tube holder, while applying moderate pressure to the tube ensuring an up/down movement in each of the holes (of the tube holder) when pulled rapidly and horizontally over 25 the top to the holder. This procedure results in the Eppendorf tube moving up and down approx. 5 mm at a frequency of 27 vertical movements per second.

The procedure is repeated 5 times over a total period of 3 seconds.

**Table 2. Sample parameters**

Sample A	Sample B
20 $\mu$ l Phi29-amplification mixture (3ng)	20 $\mu$ l Phi29-amplification mixture (3ng)

40 µl PicoSurf (2%)	40 µl PicoSurf (2%)
Vortex (1600 rpm) – 60 sec	Hand-pulled on Eppendorf holder

#### Template and droplets

Average sizes of the resulting droplets were measured using microscope images and associated measuring software. Based on the average droplet sizes the total

5 number of droplets was calculated from each sample. Based on an estimated template fragment size of 30Kb, a template amount of 3ng DNA will be equivalent to  $9.3 \times 10^7$  fragments of DNA. Thus, the preparation of both samples A and B resulted in droplets where DNA was present (details in Table ) in all created droplets.

10 The droplets generated for both samples were then incubated for 6 hours at 30°C and the reactions were terminated by raising the temperature to 65°C for 10 minutes. The content of the droplets was extracted by applying PicoBreak as described by the manufacturer. SSO Advance Supermix (BioRad) was used to quantify the target region (Dusp3) using RealTime PCR quantification (Ct). Total

15 DNA after amplification was measured using a TapeStation (Agilent).

#### Results

A collection of results are presented in Table . Briefly, the data presented in the table shows that the average measured diameter correlated to the intensity of the vortex applied. Sample A, where the most gentle handling was applied

20 resulted in an average droplet diameter of 45.7 µm whereas the hardest physical vortex (Sample B) resulted in droplets with an average diameter of 13.3 µm.

Upon Phi amplification and product extraction, DNA-quantification showed more than double quantities of amplified DNA from Sample A (28.5 ng/µl) compared to sample B (13.2 ng/µl). Regardless of the surplus of template in both samples

25 (template is present in all droplets) an approximately two-fold lower level of total DNA was observed in the reaction performed in the smaller droplets. If inhibition had not occurred, comparable DNA amounts would be expected as the total reaction volumes of both samples are identical and template is present in all droplets. Ct values in the product from the two reactions were similar. Although

this indicates that some inhibition may occur, significant amounts are still produced from the amplification in the smaller droplets and subsequent use in downstream sequencing is still possible.

5 The results obtained here suggest that droplet size (ranging from 1.2 to 50 pL) is of importance for amplification yields but that the Phi29 polymerase is relatively insensitive to the inhibition and that the inhibition effect does not prevent downstream use in sequencing.

**Table 3 Results from Droplet Phi\***

	Avg diameter ( $\mu$ m)	Droplets per 20 pL	DNA fragments pr. droplet	Phi29 amplification prod (ng/ $\mu$ L)	Ct
Sample A	45,7	4,01E+05	2,32E+02	28,5	27,29
Sample B	13,3	1,61E+07	5,79E+00	13,2	27,86

\*The table shows data from amplification in droplets. The average diameter and the number of droplets 10 is estimated based on microscopy, DNA fragments per droplet is based on Tapestation measurement of the DNA prior to droplet generation and the average droplet size, the Phi29 polymerase amplification product is determined based on Tapestation measurements.

### **Sequence listing**

MB8 Fw1 tm60: GACGGTAGATTGAGGTAATGC [SEQ ID No: 1]

15 MB8 Rev1 tm60: TATGGCCGGCGTATTAGAAG [SEQ ID No: 2]

MB10.1fw: TCAACAAACCTCGCATCGG [SEQ ID No: 3]

MB10.1rev: GTGCTGGCTGCCTGTTAC [SEQ ID No: 4]

HPV18 104 fw1: CAGATCCTTATGGGGATTCC [SEQ ID No: 5]

HPV18 105 rev1: GATTGAGGCACAGTGTAC [SEQ ID No: 6]

20 HPV18 106p: GGCATTTTGGAAATAGGGCAGG [SEQ ID No: 7]

HPV5901f: GTTAGTGTGGGCCTGTGC [SEQ ID No: 8]

HPV5994r: GGCATGGAACTTCAGTGT [SEQ ID No: 9]

ThrB-Fw: ACATTCACTCTAACAC [SEQ ID No: 10]

ThrB-Re: AATTGCTTACCCAGTTC [SEQ ID No: 11]

Beacon-ThrB: CGCGATCGTTCTGACGGCAGCTTATCGGATCGCG [SEQ ID No: 12]

E. coli amplicon1 (ThrA)

GACGGTAGATTGAGGTAATGCCCACTGCCAGCAGTTTCGACCGGATCGATAACA

5 GTAACGTTGTGACCGCGCGCTTCTAATACGCCGGCCATA [SEQ ID No: 13]

E. coli amplicon2 (ThrB)

ACATTCACTCAACAACCTCGGACGCTTGCCGATAAGCTGCCGTAGAACACCACGGGA  
AAATATCGTTATCAGTGCTGGAGCGTTTGCCAGGAACGGTAAGCAAATT [SEQ  
ID No: 14]

10 K-ras amplicon

TAGTGTATTAACCTTATGTGTGACATGTTCTAATATAGTCACATTTCTTATTATTTATTAT  
AAGGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGATGGCGTAGG  
CAAGAGTGCCTTGACGATAACAGCTAATTAGAATCATTGTGGACGAATATGATCCAA  
CAATAGAGGTAA [SEQ ID No: 15]

15 HPV18 amplicon

GATTGAGGCACAGTGTCAACCATAGTACCTGCCCTATTCCAAAATGCCTAGCAAAAG  
CTGCTCACGCCGTAAGCAAAAAACATGGAATCCCCATAAGGATCTG [SEQ ID No:  
16]

C-Thr 1325f: CCCGCGCCAATATCAACA [SEQ ID No: 17]

20 C-Thr 1485r: ACCGACGCCAATCACAAACA [SEQ ID No: 18]

## REFERENCES

1. Tan, Y-C., et al., *Monodispersed microfluidic droplet generation by shear focusing microfluidic device*. Sensors and Actuators, 2006, B114: 350-356
2. Sharma, S., et al., *Droplet-based microfluidics*. Methods in Molecular Biology, 2013. 25 949: p. 207-30.
3. Hindson, B.J., et al., *High-throughput droplet digital PCR system for absolute quantitation of DNA copy number*. Anal Chem, 2011. 83(22): p. 8604-10.
4. Sambrook, J. and D.W. Russell, *Molecular Cloning a laboratory manual*. 2001: Cold Spring Harbor Laboratory Press.
- 30 5. Walter, N.G., *Single molecule tools: fluorescence based approaches, part A. Preface*. Methods in Enzymology, 2010. 472: p. xxi-xxii.
6. Kintses, B., et al., *Microfluidic droplets: new integrated workflows for biological experiments*. Current Opinion in Chemical Biology, 2010. 14(5): p. 548-55.

7. Rinke, C., et al., *Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics*. Nat. Protocols, 2014. **9**(5): p. 1038-1048.

5 8. Agresti, J.J., et al., *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution*. Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4004-9.

9. Longo, M.C., M.S. Berninger, and J.L. Hartley, *Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions*. Gene, 1990. **93**(1): p. 125-8.

10. Prodělalová, J., et al., *Isolation of genomic DNA using magnetic cobalt ferrite and silica particles*. Journal of Chromatography A, 2004. **1056**(1-2): p. 43-48.

11. Pandit, K.R., et al., Assessment of surfactants for efficient droplet PCR in mineral oil using the pendant drop technique. *Colloids Surf B Biointerfaces*, 2015. **126**: p. 489-95.

12. Liu, Y., S.-Y. Jung, and C.P. Collier, Shear-Driven Redistribution of Surfactant Affects Enzyme Activity in Well-Mixed Femtoliter Droplets. *Analytical Chemistry*, 2009. **81**(12): p. 4922-4928.

15

## CLAIMS:

1. An *in vitro* method for enriching for one or more target DNA molecule from a sample of mixed DNA molecules comprising the steps of:
  - 5 a) providing a liquid sample of mixed DNA molecules comprising one or more specific target DNA molecule and reagents for specific detection of at least one of said target DNA molecules (401),
  - b) formation of a multiple of liquid droplets from said liquid sample (403),
  - c) specific detection of droplets containing at least one of said target 10 DNA molecules (404), wherein each droplet contains less than 0.5, preferably less than 0.25 or even more preferably less than 0.1 of said one of more target DNA molecule on average (404), and
  - d) physically selecting droplets containing at least one of said target DNA molecules (405),
  - 15 e) general amplification of DNA molecules in the coalesced selected droplets from step d), where the DNA molecules are in a total amount of less than 300fg, wherein general amplification reagents are added to the coalesced selected droplets to form a general amplification reaction mixture, and wherein from at least  $1.2 \times 10^6$  and up to a 20 maximum of  $1.2 \times 10^9$  droplets are formed for each 5 $\mu$ l of the reaction mixture.
2. An *in vitro* method for enriching for one or more target DNA molecule from a sample of mixed DNA molecules comprising the steps of:
  - 25 a) providing a liquid sample of mixed DNA molecules comprising one or more specific target DNA molecule and reagents for specific detection of at least one of said target DNA molecules (401),
  - b) formation of a multiple of liquid droplets each comprising mixed DNA molecules from said liquid sample (403),
  - c) specific detection of droplets containing at least one of said target 30 DNA molecules (404), wherein each droplet contains less than 0.5, preferably less than 0.25 or even more preferably less than 0.1 of said one of more target DNA molecule on average (404), and
  - d) physically selecting droplets containing at least one of said target DNA molecules (405), wherein the frequency of the target DNA 35

molecule compared to its frequency in the sample of mixed DNA molecules in step (a) is increased between  $0.01 \times (\text{total number of droplets containing DNA molecules}) \times (\text{number of droplets with target DNA})^{-1}$  and  $100 \times (\text{total number of droplets containing DNA molecules}) \times (\text{number of droplets with target DNA})^{-1}$ ,

5 e) general amplification of the mixed DNA molecules in the coalesced selected droplets from step d), where the mixed DNA molecules are in a total amount of less than 300fg, wherein general amplification reagents are added to the coalesced selected droplets to form a general amplification reaction mixture, and wherein from at least  $1.2 \times 10^6$  and up to a maximum of  $1.2 \times 10^9$  droplets are formed for each  $5\mu\text{l}$  of the reaction mixture.

10

15 3. The method according to claim 1 or 2, wherein the total number of droplets in step b) is at least  $5 \times 10^5$ .

20 4. The method according to claim 1 or 2, wherein the reagents are PCR reagents (401) and specific detection of said one or more target DNA molecule is performed by PCR (404).

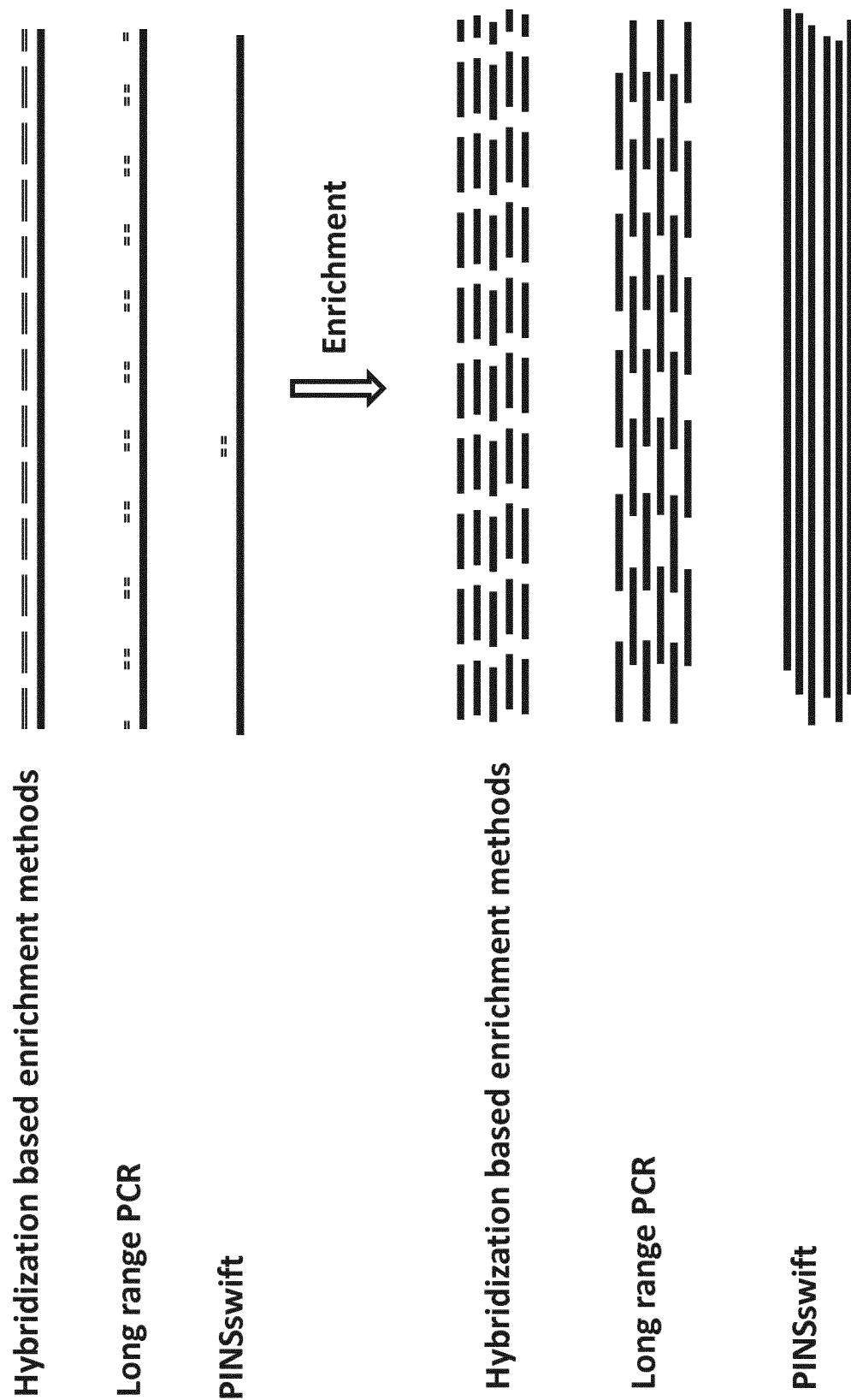
25 5. The method according to any one of claims 1-4, wherein said target DNA molecule comprises one or more unique consecutive sequence of at least 40 nucleotides.

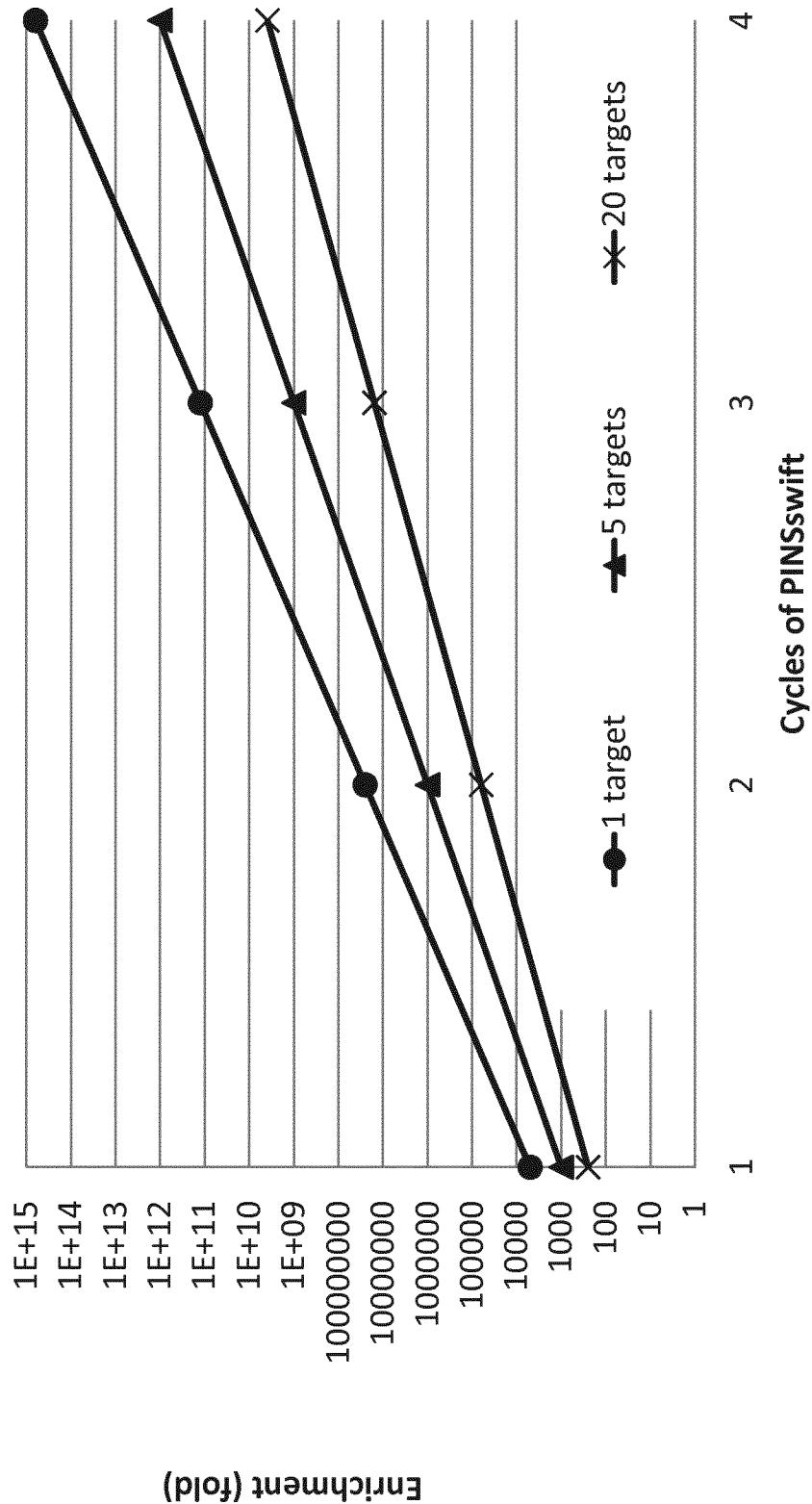
30 6. The method according to any one of claims 1-5, wherein the general amplification of DNA in step (e) is performed by Multiple Displacement Amplification.

35 7. The method according to any one of claims 1-6, wherein the reagents for specific detection contain dUTP.

8. The method according to claim 7, wherein step (d) further comprises the step of inactivating, degrading or removing DNA produced for specific detection of said one or more target DNA molecule.

9. The method according to claim 8, wherein the step of inactivating is performed using uracil-DNA N-glycosylase.
10. The method according to any one of claims 1-9, further comprising a step  
5 (f) of repeating steps (a) to (e), wherein the DNA molecules in said liquid sample of repeated step (a) are derived from the general amplification products generated from the DNA molecules in the coalesced droplets selected in step (e).
- 10 11. The method according to any one of claims 1-10, wherein each of the droplets formed in step b) have an average diameter from 2 to 20  $\mu\text{m}$  .
12. The method according to any one of claims 1-11, wherein the droplets formed in step b) contain on average at least  $10^{-7}$  target DNA molecule per  
15 droplet, preferably at least  $10^{-6}$  target DNA molecule per droplet, or even more preferably from  $10^{-4}$  to  $10^{-5}$  target DNA molecule per droplet.
13. The method according to any one of claims 1-12, wherein each of said one of more target DNA molecule of step a) comprises 1,000 to  $10^8$  nucleic acid  
20 base pairs, preferably 2,000 to 200,000 nucleic acid base pairs, more preferably 10,000 to 100,000 nucleic acid base pairs.
14. The method according to any one of claims 1-13, wherein the general amplification product generated in step e) comprises one or more amplified  
25 target DNA molecule, wherein each amplified target DNA molecule comprises 500 to 150,000 nucleic acid base pairs, preferably 2,000 to 80,000 nucleic acid base pairs, more preferably 5,000 to 50,000 nucleic acid base pairs.
- 30 15. The method according to any one of claims 1 to 14, wherein the target DNA molecule is derived from the genome of a cell.

**Figure 1**

**Figure 2**

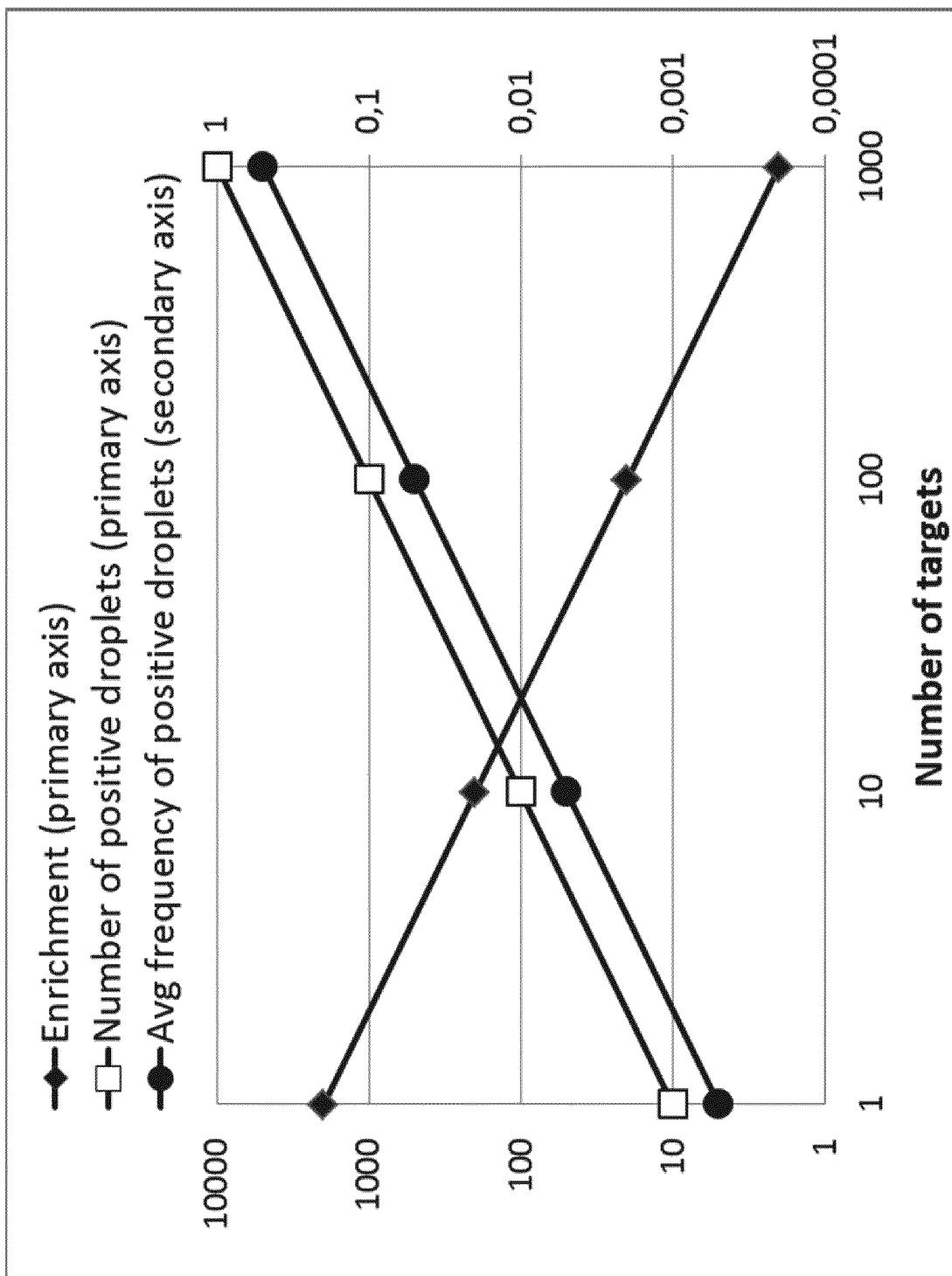
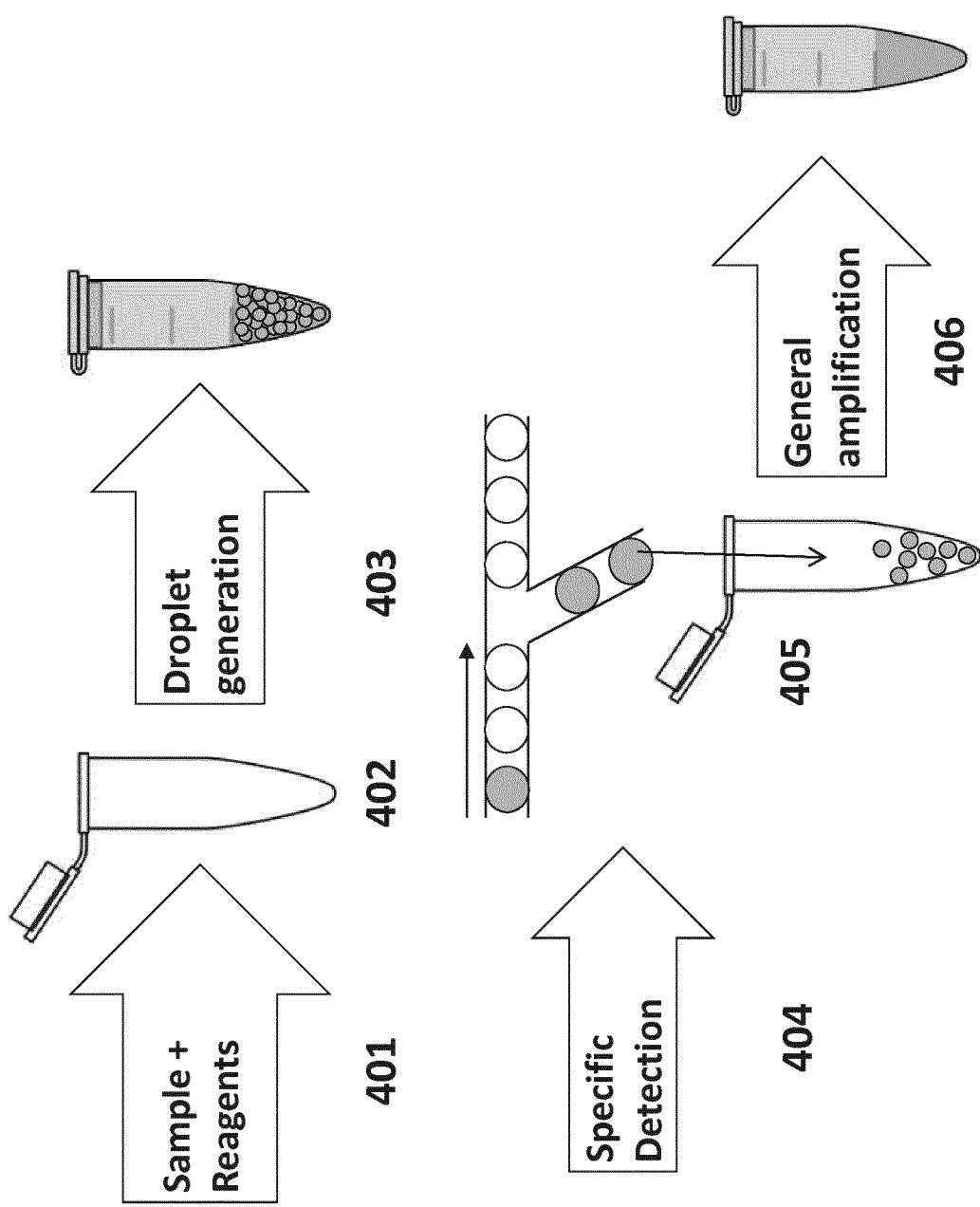


Figure 3

Figure 4



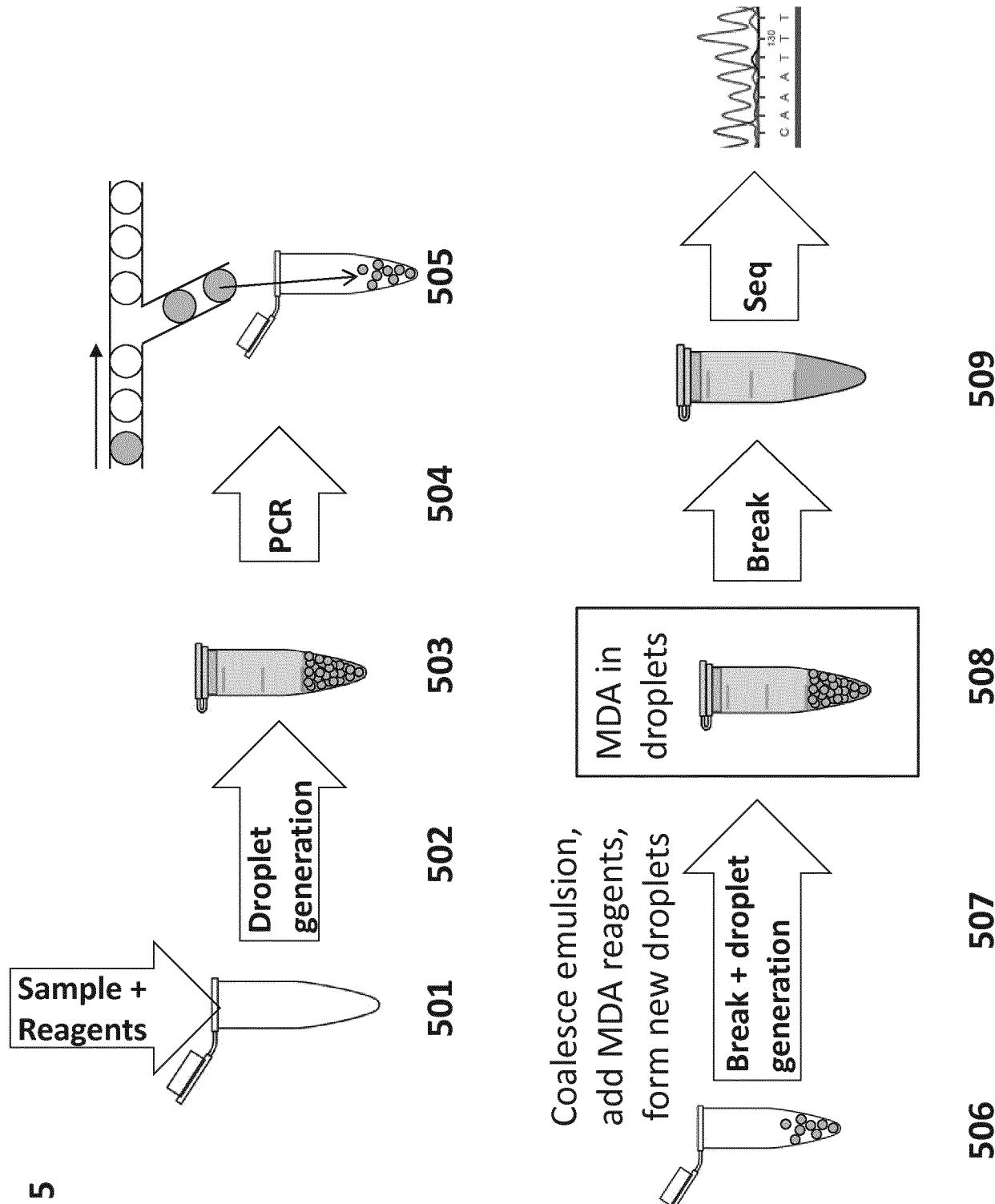
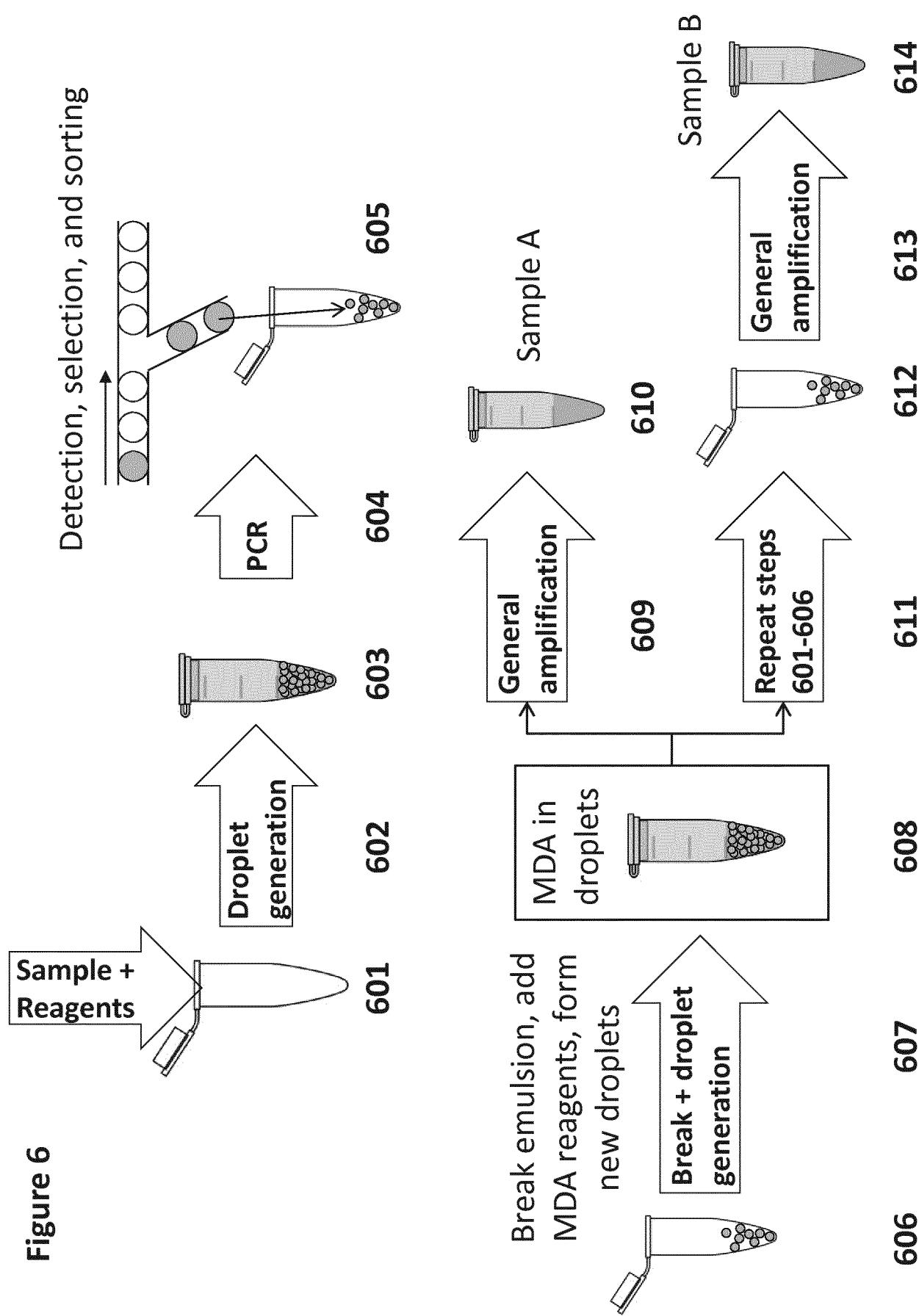
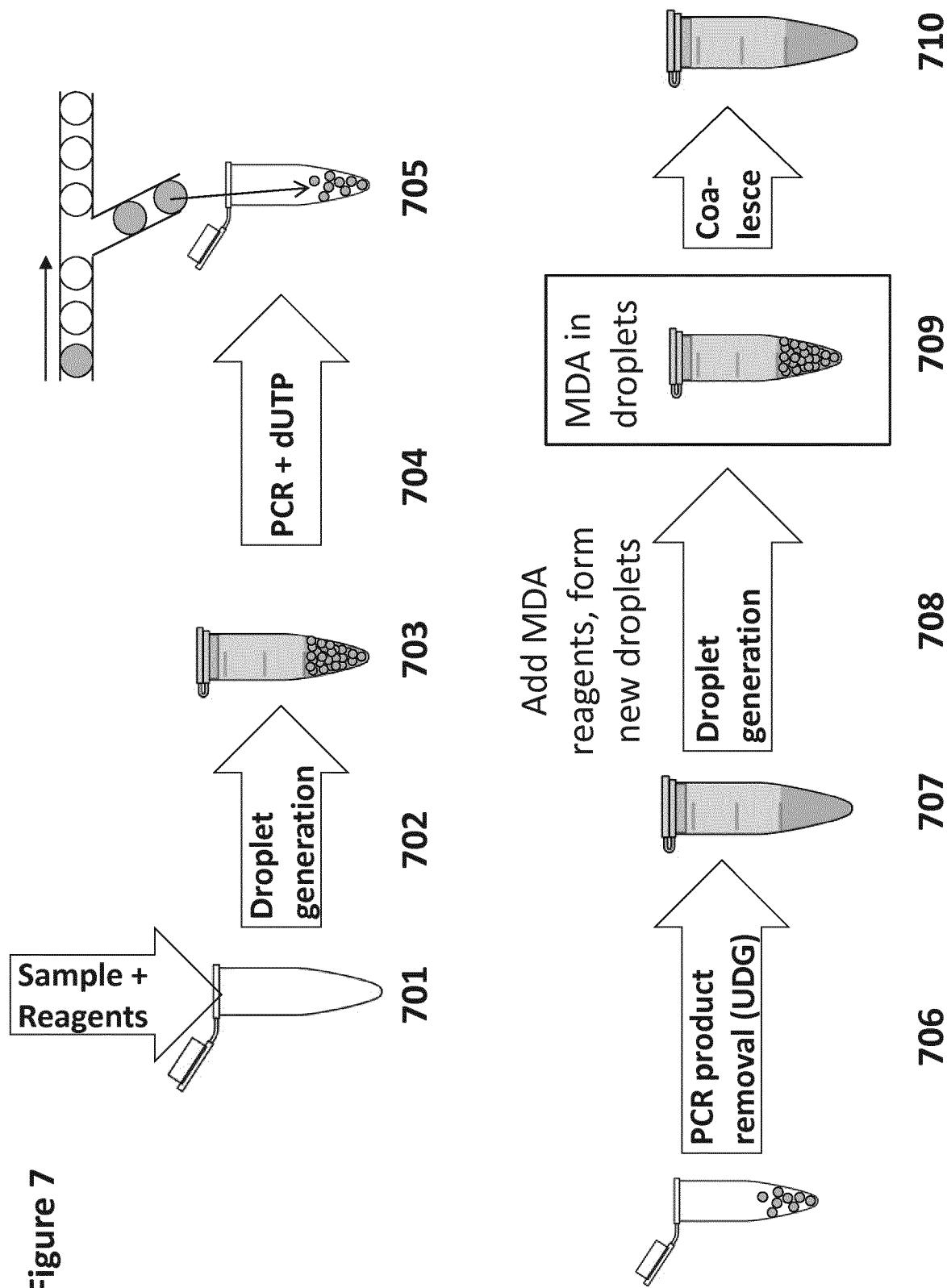
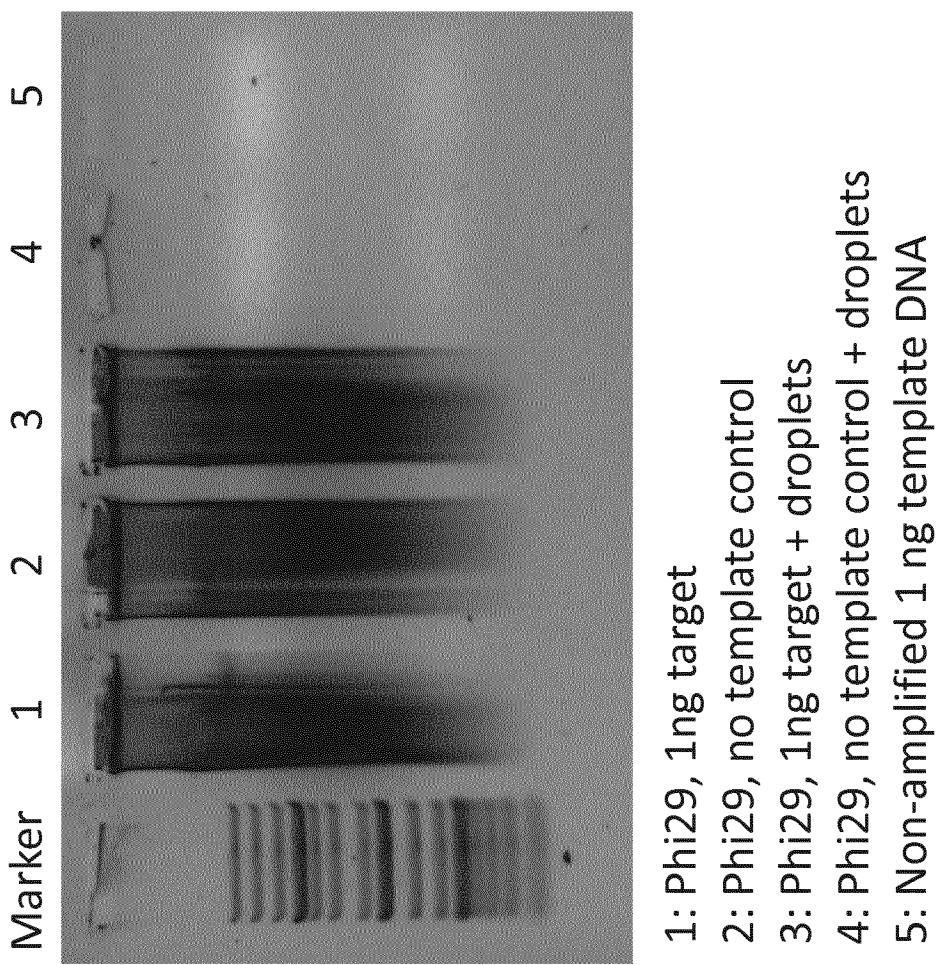


Figure 6





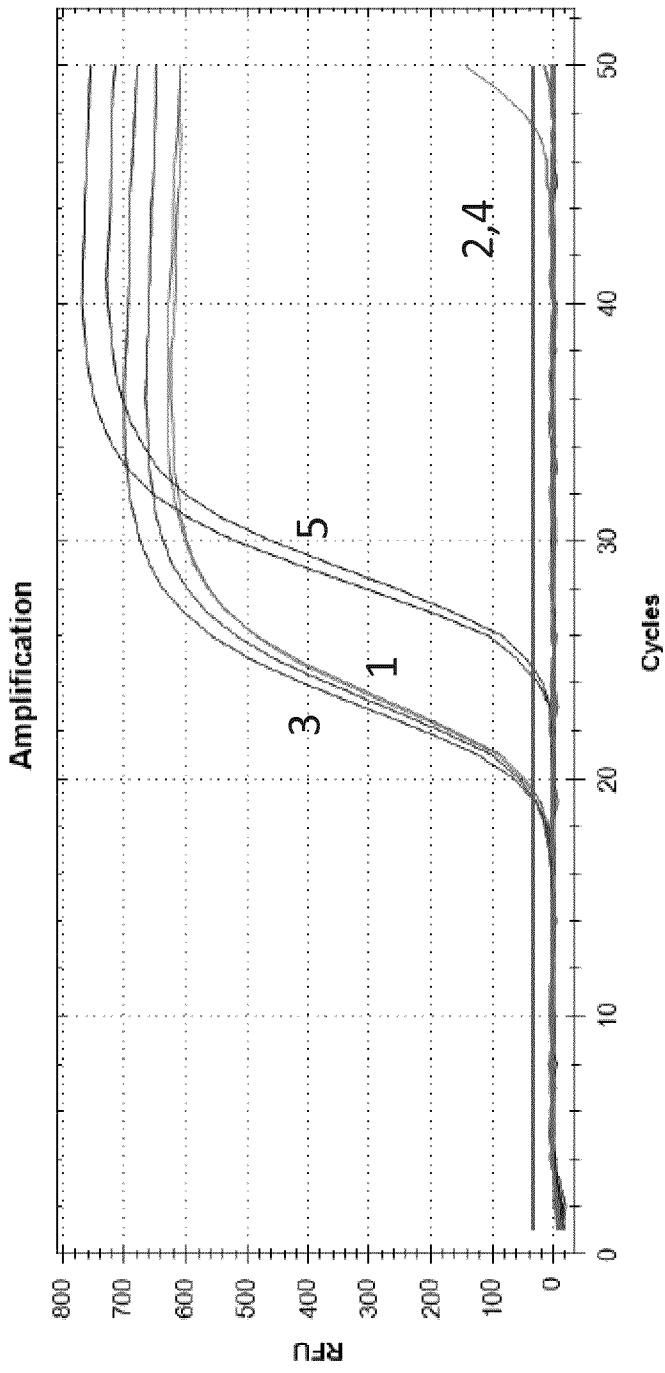
**Figure 8**

**Figure 9**

Well	Fluor	Target	Content	Sample	Cq
B01	SYBR	pos	Unkn	Single cell	19.45
B02	SYBR	pos	Unkn	Single cell	19.38
B03	SYBR	Neg	Unkn	Single cell	
B04	SYBR	Neg	Unkn	Single cell	
B05	SYBR	pos	Unkn	droplets sc	19.09
B06	SYBR	pos	Unkn	droplets sc	19.14
B07	SYBR	Neg	Unkn	droplets sc	
B08	SYBR	Neg	Unkn	droplets sc	
B09	SYBR	1E-4	Unkn	kontrol	24.22
B10	SYBR	1E-4	Unkn	kontrol	24.58

**Templates for qPCR:**

- 1: Phi29, 1ng target (B01-2)
- 2: Phi29, no template control (B03-4)
- 3: Phi29, 1ng target + droplets (B05-6)
- 4: Phi29, no template control + droplets (B07-8)
- 5: Positive control (B09-10)



# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2016/064719

**A. CLASSIFICATION OF SUBJECT MATTER**  
INV. C12Q1/68  
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DENNIS J. EASTBURN ET AL: "Microfluidic droplet enrichment for targeted sequencing", NUCLEIC ACIDS RESEARCH, vol. 43, no. 13, 14 April 2015 (2015-04-14), pages e86-e86, XP055211990, ISSN: 0305-1048, DOI: 10.1093/nar/gkv297 figure 1 -----	1-15
X, P	WO 2015/189336 A1 (SAMPLIX S A R L [LU]) 17 December 2015 (2015-12-17) claim 1 -----	1-15
X, P	WO 2015/200717 A2 (UNIV CALIFORNIA [US]; ABATE ADAM R [US]; EASTBURN DENNIS JAY [US]; SCI) 30 December 2015 (2015-12-30) paragraph [0342]; examples 9,10 -----	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 September 2016	27/09/2016

Name and mailing address of the ISA/  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3046

Authorized officer

Aslund, Fredrik

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No  
PCT/EP2016/064719

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015189336	A1	17-12-2015	NONE
WO 2015200717	A2	30-12-2015	NONE