WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 3, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced (*in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags*, Para. [0046]; *in some embodiments, the host tag comprises nine nucleotides*, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; *a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence*, Para. [0092]), but modified Steemers fails to explicitly disclose wherein the indexing site is sequenced.
However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches a indexing site is sequenced (*The preparation of multiplexed (barcoded) libraries… suitable for sequencing on the Illumina HiSeq*,Pg. 1, first full paragraph; *an adaptor oligo…(referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), the Illumina sequencing primer site*, Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include a indexing site is sequenced as taught by Meyers as multiplexed (barcoded) libraries can be sequenced with an index read primer binding site to produce efficient sequence reads.
Furthermore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have the identifier site sequenced with the indexing site, since the use of a one piece sequence construction instead of the separate sequence piece disclosed in the prior art would have required only routine skill in the art. The motivation for doing so would be to ensure that both regions are equally represented in sequence reads.

Regarding claim 4, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is sequenced (*in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags*, Para. [0046]; *in some embodiments, the host tag comprises nine nucleotides*, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; *a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence*, Para. [0092]), but modified Steemers fails to explicitly disclose wherein the indexing site is sequenced.
However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches a indexing site is sequenced (*The preparation of multiplexed (barcoded) libraries… suitable for sequencing on the Illumina HiSeq*,Pg. 1, first full paragraph; *an adaptor oligo…(referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), the Illumina sequencing primer site*, Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include a indexing site is sequenced as taught by Meyers as multiplexed (barcoded) libraries can be sequenced with an index read primer binding site to produce efficient sequence reads.
Furthermore, it would have been obvious to one of ordinary skill in the art at the time of the invention to have the identifier site sequenced separately from the indexing site, since separating sequences that were once integral involves only routine skill in the art. The motivation for doing so would be to ensure that both regions are separated during sequence reading.

Regarding claim 5, modified Steemers discloses the method of claim 1, modified Stemmers further discloses wherein the identifier site is sequenced with the target sequence (*in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags*, Para. [0046]; *in some embodiments, the host tag comprises nine nucleotides*, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; *a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence*, Para. [0092]; *methods described herein can be advantageously carried out in multiplex formats*, Para. [0149]).

Regarding claim 6, modified Steemers discloses the method of claim 1, modified Stemmers further discloses wherein the identifier site is sequenced from the target sequence (*in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags*, Para. [0046]; *in some embodiments, the host tag comprises nine nucleotides*, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; *a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence*, Para. [0092]; *methods described herein can be advantageously carried out in multiplex formats*, Para. [0149]), but modified Stemmers fails to explicitly disclose wherein the identifier site is sequenced separately from the target sequence.
It would have been obvious to one of ordinary skill in the art at the time of the invention to have an the identifier site sequenced separately from the target sequence, since separating parts that were once integral involves only routine skill in the art. The motivation for doing so would be to ensure that each sequence region alone is most efficiently amplified to optimize sequence reads.

Regarding claim 9, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the plurality of nucleic acid fragments is generated from more than one sample (*The sub-sampled target DNA; adaptors are ligated to…ends of each nucleic acid*, Para. [0167]; *steps include…adaptor ligation, and PCR amplify…and sequence products*, Para. [0080]; *some embodiments include populations of genomic DNA fragments*, Para. [0058]).

IDT Exhibit 1018.0631

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 13, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments (in some embodiments include populations of genomic DNA fragments, Para. [0058]).

Regarding claim 14, modified Steemers discloses the method of claim 13, modified Steemers further discloses wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments (in some embodiments include populations of genomic DNA fragments, Para. [0058]).

Regarding claim 15, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing site is between 2 and 8 nucleotides in length.
However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches an indexing site is 30 nucleotides in length (An adaptor oligo…(referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing site is 30 nucleotides in length as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences (with an index read primer binding site) may efficiently be used to generate sequence reads (Myers, Pg. 1, second full paragraph).
Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have an indexing site between 2 and 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that an indexing site between 2 and 8 nucleotides could be identified in the sequence reads.

Regarding claim 16, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing site is about 8 nucleotides in length.
However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches an indexing site is 30 nucleotides in length (An adaptor oligo…(referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing site is 30 nucleotides in length as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences (with an index read primer binding site) may efficiently be used to generate sequence reads (Myers, Pg. 1, second full paragraph).
Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have an indexing site about 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that an indexing site about 8 nucleotides in length could be identified in the sequence reads.

Regarding claim 17, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is 8 nucleotides in length (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]), but modified Steemers fails to explicitly disclose wherein the identifier site is between 1 and 8 nucleotides in length.
It would have been obvious to one of ordinary skill in the art at the time the invention was made to have an identifier site between 1 and 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that at least 1 to 8 nucleotides of the identifier site could be identified in the sequence reads.

Regarding claim 18, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the identifier site is about 8 nucleotides in length (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]), but modified Steemers fails to explicitly disclose wherein the identifier site is about 8 nucleotides in length.
It would have been obvious to one of ordinary skill in the art at the time the invention was made to have an identifier site about 8 nucleotides in length, since where the general conditions of the claim are disclosed in the prior art, discovering the optimum or workable ranges involves only routine skill in the art. The motivation for doing so would be to ensure that at least 8 nucleotides of the identifier site could be identified in the sequence reads.

IDT Exhibit 1018.0632

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065930

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 19, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the indexing primer binding site is a universal indexing primer binding site.
However, Myers is in the field of adaptor oligo, Pg. 1, second full paragraph and teaches an indexing primer binding site is a universal indexing primer binding site (An adaptor oligo ...(referred to as the 5' adaptor) is ligated to the 5' end; PCR primers have long tails (~30 nt) that extend the length of the product; the tails contain the barcoding sequences (with an index read primer binding site), Pg. 1, second full paragraph; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; an indexing primer binding site is a binding site for a universal indexing primer, i.e., the indexing primer binding site is a universal indexing primer binding site, Para. [0035]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an indexing primer binding site is a universal indexing primer binding site as taught by Myers as adaptors ligated on the 5' end and PCR primers containing barcoding sequences with an index read primer binding site which can be a universal site, may efficiently be used to generate sequence reads (Myers, Pg. 1, second full paragraph).

Claim 2 lacks an inventive step under PCT Article 33(3) as being obvious over Steemers et al. (hereafter Steemers) in view of Meyers (hereafter Meyers) and ILLUMINA, INC. (hereafter ILLUMINA, INC.).

Regarding claim 2, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the method further comprises the population of sequence reads the sequencing read with a duplicate identifier site and target sequence (In some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; In some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but Steemers fails to explicitly disclose wherein removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
However, ILLUMINA, INC. is in the field of removing duplicates sequences, Pg. 42, Lns. 1-3 and teaches removing from the population of sequence reads with a duplicate identifier site and target sequence (In some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Pg. 7, Lns. 1-3; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Pg. 18, Lns. 8-10; removing duplicates and counting the original representation of molecules, each having a unique tag; thus, the cDNAs can be amplified and sequenced, Pg. 42, Lns. 1-3).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include removing from the population of sequence reads with a duplicate identifier site and target sequence as taught by ILLUMINA, INC. as removing duplicates and counting the original representation of molecules, each having a unique tag enables nucleic acids can to be amplified and sequenced from various populations of sequence reads (ILLUMINA, INC., Pg. 7, Lns. 1-3, Pg. 18, Lns. 8-10, and Pg. 42, Lns. 1-3).

Claims 7, 8, 10-12, and 20 lack an inventive step under PCT Article 33(3) as being obvious over Steemers et al. (hereafter Steemers) in view of Meyers (hereafter Meyers) and SOLEXA LIMITED (hereafter SOLEXA LIMITED).

Regarding claim 7, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the adaptor comprises (iii) the identifier site; and (iv) the target sequence primer binding site (In some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; In some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but modified Steemers fails to explicitly disclose wherein the adaptor comprises from 5' to 3': (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site.
However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches an adaptor comprises from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site (The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 28; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; the tags are on the adaptors, pg. 20, Ln. 10; ligating identical adaptor polynucleotide molecules... to the 5' ... ends of the target polynucleotide, Pg. 20, Lns. 19-20 and 23-24).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an adaptor comprising from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site as taught by SOLEXA LIMITED as adaptors containing (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site in a 5 to 3 orientation can be most efficient in detecting duplicate sequences.
Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have the adaptor comprising from 5' to 3 ": (i) the indexing primer binding site; (ii) the indexing site; (iii) the identifier site; and (iv) the target sequence primer binding site, since rearranging adaptor parts of an invention involves only routine skill in the art. The motivation for doing so would be to ensure that the 5' to 3' orientation of sequences are most efficient in the determining duplication reads.

IDT Exhibit 1018.0633

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065550

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 9, modified Steemers discloses the method of claim 1, modified Steemers further discloses wherein the adaptor comprises (iii) the target sequence primer binding site; and (iv) the identifier site (In some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]; in some embodiments, the host tag comprises nine nucleotides, Para. [0064]; wherein according to the instant application, 'An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads:', Para. [0030]; a sequencing read can be carried out on a single target nucleic acid molecule or simultaneously on a population of target nucleic acid molecules having the same sequence, Para. [0092]; methods described herein can be advantageously carried out in multiplex formats, Para. [0149]), but modified Steemers fails to explicitly disclose wherein the adaptor comprises from 5' to 3 ': (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site.
However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches an adaptor comprises from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site (The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; the tags are on the adaptors, Pg. 20, Ln. 10; ligating identical adaptor polynucleotide molecules... to the 5' ...ends of the target polynucleotide, Pg. 20, Lns. 19-20 and 23-24).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include an adaptor comprising from 5' to 3' (i) the indexing primer binding site; (ii) the indexing site as taught by SOLEXA LIMITED as adaptors containing (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site in a 5 to 3 orientation can be most efficient in detecting duplicate sequences.
Furthermore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have the adaptor comprising from 5' to 3 ': (i) the indexing primer binding site; (ii) the indexing site; (iii) the target sequence primer binding site; and (iv) the identifier site, since rearranging adaptor parts of an invention involves only routine skill in the art. The motivation for doing so would be to ensure that the 5' to 3' orientation of sequences are most efficient in the determining duplication reads.

Regarding claim 10, modified Steemers discloses the method of claim 9, but modified Steemers fails to explicitly disclose wherein the nucleic acid fragments from each sample has the same indexing site.
However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches a nucleic acid fragments from each sample has the same indexing site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include nucleic acid fragments from each sample having the same indexing site as taught by SOLEXA LIMITED as samples having the same indexing site can be sequenced together, eliminating the need for multiple sequencing runs.

Regarding claim 11, modified Steemers discloses the method of claim 10, but modified Steemers fails to explicitly disclose wherein the sequencing reads are separated based on the indexing site.
However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches sequencing reads are separated based on the indexing site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; the tag and target are read in a single read, Pg. 13, Lns. 29-30; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0029]; wherein said sequencing is performed in two separate reads, Pg. 97, Claim 4.
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include sequencing reads are separated based on the indexing site as taught by SOLEXA LIMITED as duplicate sequence reads in multiple samples can be determined when sequence reads are separated based on tags and/or identifiers.

Regarding claim 12, modified Steemers discloses the method of claim 11, modified Steemers further discloses wherein the separation of sequencing reads is performed (in some embodiments, the assembling step further comprises identifying more than one sequencing read comprising the same host tags, Para. [0046]), but modified Steemers discloses fails to explicitly disclose wherein the separation of sequencing reads is performed prior to step c).
It would have been obvious to one of ordinary skill in the art at the time the invention was made to have the separation of sequencing reads is performed prior to step c), since selection of any order of process steps which did not result in a new or unexpected result would involve only routine skill in the art. The motivation for doing so would be to ensure the duplicate sequence reads are identified accurately.

Form PCT/ISA/237 (Supplemental Box) (July 2011)

IDT Exhibit 1018.0634

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US2014/065530

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

Regarding claim 20, modified Steemers discloses the method of claim 1, but modified Steemers fails to explicitly disclose wherein the target sequence primer binding site is a universal target sequence primer binding site.
However, SOLEXA LIMITED is in the field of indexing samples (Title) and teaches a target sequence primer binding site is a universal target sequence primer binding site (The universal forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct, Pg. 4, Lns. 5-8; provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends; target sequence can therefore be any DNA fragment of interest Pg. 4, Lns. 13-17; each DNA sample, Pg. 4, Ln. 25; the tag and target are read in a single read, Pg. 13, Lns. 29-30; wherein according to the instant application, 'An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site', Para. [0039]).
It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the method of Steemers to include a target sequence primer binding site is a universal target sequence primer binding site as taught by SOLEXA LIMITED as samples having the same universal target sequence primer binding site can be sequenced together, eliminating the need for multiple sequencing runs.

Claims 1-21 meet the criteria set out in PCT Article 33(4), and thus have industrial applicability because the subject matter claimed can be made or used in industry.

IDT Exhibit 1018.0635

Search History:

Limited Classification Search

The Patent Analyst performed a limited classification search within the following US, IPC, CPC, ECLA, or F-Term classification areas:

U.S. Class/Subclass(es): 435/6.1, 6.11; 506/2, 23; 536/23.1, 24.2, 25.3

IPC(8) Class/Subclass(es): C12N 15/09, 15/10, 15/11; C12P 19/34, C12Q 1/68; C40B 20/00, 20/04 *(2015.01)*

CPC Class/Subclass(es): C12Q 1/6869, 1/6874, 2525/16, 2537/143

See Global Search Results.

Global Patent Literature Text Search

The Patent Analyst performed the following global text search, which was not limited by classification but may or may not have been limited by other criteria:

PATBASE

#	SEARCH	Results
1	PA=("NUGEN_TECHNOLOGIES" OR "NUGEN_TECHNOLOGIES_INC" OR "NUGEN_TECHNOLOGIES_INC201_INDUSTRIAL_RD_SAN_CARLOSCA94070US" OR "NUGEN_TECHNOLOGIES_INC821_INDUSTRIAL_ROADSAN_CARLOSCA95050US")	28
2	INV=("AMORESE_DOUGLAS" OR "AMORESE_DOUGLAS_A" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707_US" OR "AMORESE_DOUGLAS_ALAN")	38
3	INV=("AMORESE_DOUGLAS" OR "AMORESE_DOUGLAS_A" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707" OR "AMORESE_DOUGLAS_A_HOCKESSIN_DELAWARE_19707_US" OR "AMORESE_DOUGLAS_ALAN")	38
4	INV=("SCOLNICK_JONATHAN")	4
5	INV=("SCOLNICK_JONATHAN")	4
6	INV=("SCHROEDER_BEN")	1
7	INV=("SCHROEDER_BENJAMIN" OR "SCHROEDER_BENJAMIN_G" OR "SCHROEDER_BENJAMIN_R"	18

	OR "SCHROEDER_BENJAMIN_V")	
8	PN=US20090298075	1
9	PN=WO2011156529	1
10	PN=US20100330574	1
11	PN=WO2008106987	1
12	(C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)	1
13	PN=US20120208705	1
14	PN=US20140256571	1
15	PN=US8847799	1
16	PN=WO2008093098	1
17	FT=(nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)	142
18	FT=((nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	25
19	FT=((nucleic acid) and (barcode) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and IC=((C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	25
20	FT=((nucleic acid) and (sequenc*) AND (tag*) AND (duplicate*) AND (identifier*) AND (adaptor*)) and IC=((C12N15/09 OR C12N15/10 OR C12N15/11 OR C12P19/34 OR C12Q1/68 OR C40B20/00 OR C40B20/04)) and CPC=((C12Q1/6869 OR C12Q1/6874 OR C12Q2537/143 OR C12Q2525/16)) and UC=((435/6.1 OR 435/6.11 OR 506/2 OR 506/23 OR 536/23.1 OR 536/24.2 OR 536/25.3))	40

GOOGLE PATENT: www.google.com/advanced_patent_search

No. of Hits	Text String
6,240	Single Primer Extension (SPET) Reaction
609	indexing primer binding site
420	method of obtaining duplicate sequence reads, ligated adaptor on 5' end,

	index site
363	method of detecting duplicated sequence, ligated adaptor on 5' end, index site
5,430	Sequencing duplicate nucleic acid fragments
662	Sequencing duplicate nucleic acid fragments and ligated adaptor at 5' end
3,92	identifying a duplicate reads
1,720	identify duplicate sequence reads
517	kit comprising adaptor, index primer binding site, index site
1,110,000	a method for multiplexing duplicate sequencing read
1,020	Barcoding nucleic acids
383,000	universal indexing primer binding site

Computer Accessed Text Databases Searched

The Patent Analyst searched the following computer accessed text databases:

GOOGLE: www.google.com

Search	Query	Items found
#1	A method for detecting a duplicate sequencing read	51,200,000
#2	detecting duplicate sequence reads using multiplex and ligated adaptors	78,800
#3	indexing primer binding site and adaptors	738,000

PUBMED: www.ncbi.nlm.nih.gov/pubmed/advanced

Search	Query	Items found
#1	SCHROEDER, Ben	1
#2	SCOLNICK Jonathan[Author]	6
#3	AMORESE Douglas[Author]	0

Date search was completed: 18 February 2015

DF/JJ

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 August 2008 (07.08.2008)

PCT

(10) International Publication Number
WO 2008/093098 A2

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

(21) International Application Number:
PCT/GB2008/000337

(22) International Filing Date: 1 February 2008 (01.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/899,221 2 February 2007 (02.02.2007) US

(71) Applicant *(for all designated States except US)*: **SOLEXA LIMITED** [GB/GB]; Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BIGNALL, Helen** [GB/GB]; SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB). **FRASER, Louise** [GB/GB]; SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB). **GORMLEY, Niall** [IE/GB]; SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB).

(74) Agents: **WHITE, Nina, Louise** et al.; BOULT WADE TENNANT, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(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, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(54) Title: METHODS FOR INDEXING SAMPLES AND SEQUENCING MULTIPLE NUCLEOTIDE TEMPLATES

(57) Abstract: The invention relates to methods for indexing samples during the sequencing of polynucleotide templates, resulting in the attachment of tags specific to the source of each nucleic acid sample such that after a sequencing run, both the source and sequence of each polynucleotide can be determined. Thus, the present invention pertains to analysis of complex genomes (e.g., human genomes), as well as multiplexing less complex genomes, such as those of bacteria, viruses, mitochondria, and the like.

WO 2008/093098 A2

IDT Exhibit 1018.0639

Methods for indexing samples and sequencing multiple polynucleotide templates

Field of the invention

The invention relates to methods for indexing samples during the sequencing of polynucleotide templates, resulting in the attachment of tags specific to the source of each nucleic acid sample such that after a sequencing run, both the source and sequence of each polynucleotide can be determined. Thus, the present invention pertains to analysis of complex genomes (e.g., human genomes), as well as multiplexing less complex genomes, such as those of bacteria, viruses, mitochondria, and the like.

Background to the invention

Several publications and patent documents are referenced in this application in order to more fully describe the state of the art to which this invention pertains. The disclosure of each of these publications and documents is incorporated by reference herein.

Advances in the study of biological molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for sequence analysis.

The study of complex genomes, in particular, the search for the genetic basis of disease in humans requires genotyping on a massive scale. Screens for numerous genetic markers performed on populations large enough to yield statistically significant data are needed before associations can be made between a given genotype and a

particular disease. However large-scale genotyping is demanding in terms of the cost of both materials and labour involved, and the time taken to perform the study, especially if the methodology employed involves separate serial analysis of individual DNA samples. One shortcut is to pool DNA from many individuals and to determine parameters such as the ratio of changes at certain positions in the genome. Such measurements of 'allele frequency' in the pool of samples can be used to correlate the relationship between the changes in the genome sequence and the occurrence of a disease. Hence, an association study involving 1000 patients would in theory only necessitate a 'one-pot' reaction for each genetic change. Pooling therefore represents an effective technique for analysing large quantities of samples in a facile manner.

One disadvantage of pooling samples prior to analysis is that information pertaining to individual DNA samples is lost; only global information such as allele frequencies is gathered, as there is no easy method for discerning which individuals gave rise to a particular genotype. An ability to genotype large populations in a small number of reactions, while retaining the information relating to the source of the individual samples, would yield the information content of a full non-pooled population screen in the time and at the cost of a pooled reaction.

Several of the new methods employed for high throughput DNA sequencing (*Nature*. 437, 376-380 (2005); *Science*. 309, 5741, 1728-1732 (2005)) rely on a universal amplification reaction, whereby a DNA sample is randomly fragmented, then treated such that the ends of the different fragments all contain the same DNA sequence. Fragments with universal ends can be amplified in a single reaction with a single pair of

amplification primers. Separation of the library of fragments to the single molecule level prior to amplification ensures that the amplified molecules form discrete populations that can then be further analysed. Such separations can be performed either in emulsions (*Nature*. 437, 376-380 (2005); *Science*. 309, 5741, 1728-1732 (2005)), or on a surface (Nucleic Acids Research 27, e34 (1999); Nucleic Acids Research 15, e87 (2000)).

WO 98/44151 and WO 00/18957 both describe methods of forming polynucleotide arrays based on 'solid-phase' nucleic acid amplification, which is a bridging amplification reaction wherein the amplification products are immobilised on a solid support in order to form arrays comprised of nucleic acid clusters or 'colonies'. Each cluster or colony on such an array is formed from a plurality of identical immobilised polynucleotide strands and a plurality of identical immobilised complementary polynucleotide strands. The arrays so-formed are generally referred to herein as 'clustered arrays' and their general features will be further understood by reference to WO 98/44151 or WO 00/18957, the contents of both documents being incorporated herein in their entirety by reference.

In common with all amplification techniques, solid-phase bridging amplification requires the use of forward and reverse amplification primers which include 'template-specific' nucleotide sequences which are capable of annealing to sequences in the template to be amplified, or the complement thereof, under the conditions of the annealing steps of the amplification reaction. The sequences in the template to which the primers anneal under conditions of the amplification reaction may be referred to herein as 'primer-binding' sequences.

Certain embodiments of the methods described in WO 98/44151 and WO 00/18957 make use of 'universal' primers to amplify templates comprising a variable template portion that it is desired to amplify flanked 5' and 3' by common or 'universal' primer binding sequences. The 'universal' forward and reverse primers include sequences capable of annealing to the 'universal' primer binding sequences in the template construct. The variable template portion, or 'target' may itself be of known, unknown or partially known sequence. This approach has the advantage that it is not necessary to design a specific pair of primers for each target sequence to be amplified; the same primers can be used for amplification of different templates provided that each template is modified by addition of the same universal primer-binding sequences to its 5' and 3' ends. The variable target sequence can therefore be any DNA fragment of interest. An analogous approach can be used to amplify a mixture of templates (targets with known ends), such as a plurality or library of target nucleic acid molecules (e.g. genomic DNA fragments), using a single pair of universal forward and reverse primers, provided that each template molecule in the mixture is modified by the addition of the same universal primer-binding sequences.

DNA from more than one source can be sequenced on an array if each DNA sample is first tagged to enable its identification after it has been sequenced. Many low scale DNA-tag methodologies already exist, for example fluorescent labelling (Haughland, Handbook of Fluorescent Probes and Research Products, Invitrogen/Molecular Probes), but these are limited in scope to less than 10 or so reactions in parallel. DNA tags can be added to the ends of DNA fragments by cloning, of example as described in US 5604097. The tags

consist of eight four base 'words', where each word uses only three bases (A,T and C) in various combinations resulting in a total of 16,777,216 different tags that all have the same base pair composition and melting points. Such tags are used to label target molecules in a sample so that after an amplification reaction, each original molecule in the sample has a unique tag. The tags can then be used to 'sort' the sample onto beads containing sequences complementary to the tags such that each bead contains multiple copies of a single amplified target sequence (Brenner et al., (2000) Nature Biotechnology, 18, 630). In this application the tags are not sequenced, so the method does not provide a method of analysing targets from multiple samples, but rather a method of sorting a mixture of amplified templates from a single sample. The problem with enabling the method for individual samples rather than individual molecules is that the tags are synthesised in a combinatorial manner, meaning that all 16,777,216 different sequences are obtained in a single mixture in the same tube. Whilst this is ideal for treating one sample such that each individual molecule in the sample carries a different tag, it does not permit attachment of the same tag to every molecule in the sample.

DNA samples from multiple sources can, however, be tagged with different nucleic acid tags such that the source of the sample can be identified. Previous application WO05068656 describes the generic concept of indexing samples. In order to utilise this invention on arrays of amplified single molecule templates, for example as described in WO9844151, WO06099579 or WO04069849, it is advantageous to prepare the nucleic acids using the novel method described herein. The optimised DNA sample

preparation techniques described herein are applicable to any method where the samples are amplified prior to sequencing. The DNA sample preparation techniques presented herein describe in detail the optimal placements of the sequencing primers and indexed tags within the DNA constructs to be sequenced.

Summary of the invention

The present inventors have developed methods for indexing samples, wherein the samples are amplified from isolated single template molecules. Using the techniques of the invention, it is possible to prepare a nucleic acid sample for a sequencing reaction wherein both the target sequence and the tag sequence can be determined.

In a first embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a universal adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by universal adaptor sequences and said universal adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid;

(c) amplifying adaptor-target-adaptors of each sample with two or more sample specific amplification primers to generate amplified nucleic acids, wherein one of said amplification primers comprises a sample specific tag sequence, and wherein amplified nucleic acids of each sample comprise said sample specific tag sequence and said sample specific tag sequence differentiates amplified nucleic acids originating from different samples;

(d) pooling the amplified nucleic acids of different samples;

(e) immobilising the amplified nucleic acids of different samples on an array to generate an array of immobilised fragments;

(f) sequencing the immobilised fragments on the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In a second embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-

target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples;

(c) pooling the adaptor-target-adaptors of different samples to generate a pooled sample of adaptor-target-adaptors;

(d) amplifying the pooled sample of adaptor-target-adaptors with a pair of universal amplification primers to generate an amplified pooled sample;

(e) immobilising the amplified pooled sample on an array to generate an array of immobilised fragments;

(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In a third embodiment of the invention, a method is presented for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, wherein the method comprises the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples, and wherein the sample specific tag sequence is attached directly to the target fragment in each adaptor-target-adaptor with no intervening nucleotides between the sample specific tag sequence and target fragment sequences;

(c) amplifying adaptor-target-adaptors of each sample with a pair of amplification primers complementary to the sample specific tag sequence to generate amplified nucleic acids from each sample;

(d) pooling the amplified nucleic acids from each sample to generate a pool of amplified nucleic acid samples;

(e) immobilising the pool of amplified nucleic acid samples on an array to generate an array of immobilised fragments;

(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

In accordance with the first or second embodiment of the invention, sequencing can be performed in two separate reads. In accordance with the third embodiment of the invention, sequencing of the target fragment and the sample

specific tag sequence can be performed in a single sequencing read.

With respect to the first, second, or third embodiments, the method encompasses a step wherein the pooled samples are amplified on a solid surface prior to sequencing. In a particular aspect, the solid surface on which amplification is performed is a collection of beads. In a more particular aspect, the collection of beads may be isolated as single beads in an emulsion.

With respect to the first, second, or third embodiments, the method encompasses a step wherein the array of immobilized fragments is amplified prior to sequencing to generate an amplified array. In a particular aspect, an amplified array so produced is a clustered array of amplified single molecules. Such clustered arrays may be formed, for example, by solid-phase nucleic acid amplification or isothermal solid-phase nucleic acid amplification.

The present method encompasses sequencing that involves cycles of ligation with labelled oligonucleotides or cycles of addition of nucleotides. In a particular aspect of the invention, the nucleotides are labelled. In a more particular aspect the nucleotides are labelled with fluorophores.

In an aspect of any of the first, second, or third embodiments of the invention, the sequencing step may be carried out on an array of immobilised beads.

A further aspect of the invention comprises a kit comprising a double stranded nucleic acid adapter and at least two different primer polynucleotides that carry a different tag sequence that does not hybridise to the adapter.

Brief description of the figures

Figure 1 shows a schematic of an indexed sample construct using the first method of the invention, using universal adaptors. Common adaptors are used for all tags, tags are added using one unique amplification primer and sequencing is carried out in two reads.

Figure 2 shows a schematic of an indexed sample construct using the second method of the invention, using tagged adaptors. Tagged adaptors are used for each sample and amplified with one pair of universal primers. Sequencing is carried out in two reads.

Figure 3 shows a schematic of an indexed sample construct using the third method of the invention, using adjacent tag-targets. Two adaptors and two amplification primers are used per tag and sequencing carried out in a single read.

Figure 4 shows numerical data on two sequencing reads obtained from a sample prepared according to the first method of the invention, amplified into an array of clusters and sequenced with two sequencing reads, one for the target and one for the tag.

Figure 5 shows the concept of tagging a sample for use in a paired end reading process. Libraries are amplified at the PCR step using a nested PCR reaction with 3 primers (2 universal and 1 tag oligo.

Figure 6a shows a nested PCR reaction using three amplification primers, one of which carries the tag.

Figure 6b shows the same nested PCR concept as figure 6, but with the benefit of the sequences of the various adapters and primers.

Figure 7 shows the embodiment of the method where the first sequencing read is performed, followed by removing the first sequencing primer, hybridising a second sequencing primer to determine the tag sequence, inverting the strand on the surface using a strand resynthesis protocol, and performing a second read at the opposite end of the fragment to the first read to generate a pair of reads from each fragment along with the tag read.

Figure 8 shows some exemplary data obtained using the method shown in figure 7.

Figure 9 shows the sequence of the various oligonucleotides used in the process

Figure 10 shows a schematic of MJ Research thermocycler, coupled with an 8-way peristaltic pump Ismatec IPC ISM931 equipped with Ismatec tubing (orange/yellow, 0.51 mm ID).

Detailed description of the invention

The present invention is directed to methods that advance techniques for nucleic acid analysis, such that target nucleic acids from a plurality of sources can be tagged with a unique, identifiable sequence of nucleic acid bases and then sequenced on a single array. As described

herein, the presence of a characteristic nucleic acid sequence tag on an immobilised target molecule permits identification of the source of the target, concurrent with a sequence read from the target. This is a dramatic improvement over pre-existing array technologies which generally require a two-step process involving an initial sequencing step for sequencing the pooled nucleic acid targets, followed by a second analysis step wherein the source of the nucleic acid is determined.

The limitations of such pre-existing array technologies are evident in WO 98/44151 and WO 00/18957, for example, which both describe methods of nucleic acid amplification which allow amplification products to be immobilised on a solid support in order to form arrays comprised of clusters or "colonies" formed from a plurality of identical immobilised polynucleotide strands and a plurality of identical immobilised complementary strands. The nucleic acid molecules present in DNA colonies on the clustered arrays prepared according to these methods can provide templates for sequencing reactions, but to date only a single sequencing read can be obtained from each immobilised strand in each colony. The methods described herein allow for two or more reads on a single amplified strand, thereby allowing analysis of the tag sequence independently of the target sequence, and also allowing the possibility of a paired read from the opposite end of each fragment as well as a third read for the tag sequence. The methods also allow the preparation of constructs where the tag sequence is adjacent to the target sequence, and the tag and target are read in a single read, which again can be turned into a paired read methodology if desired.

As will be apparent to the skilled reader, references herein to a particular nucleic acid sequence may, depending on the context, also refer to nucleic acid molecules which comprise the nucleic acid sequence. Sequencing of a target fragment means that a read of the chronological order of bases is established. The bases do not need to be contiguous, although this is preferred, nor does every base on the entire fragment have to be sequenced.

The following passages describe different aspects of the invention in greater detail. Each aspect of the invention may be combined with any other aspect or aspects of the invention unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature of features indicated as being preferred or advantageous.

The terms 'target nucleic acid sequence', 'target nucleic acid molecule', 'target nucleic acid' and 'target nucleic acid fragment' may be used interchangeably to refer to nucleic acid molecules that it is desired to sequence on an array according to the invention. The target nucleic acid may be essentially any nucleic acid of known or unknown sequence. It may be, for example, a fragment of genomic DNA or cDNA. Sequencing may result in determination of the sequence of the whole, or a part of the target molecule. The targets can be derived from a primary nucleic acid sample that has been randomly fragmented. The targets can be processed into templates suitable for amplification by the placement of universal amplification sequences at the ends of each target fragment. The targets can also be obtained from a primary RNA sample by reverse transcription into cDNA.

As used herein, the term 'polynucleotide' refers to deoxyribonucleic acid (DNA), but where appropriate the skilled artisan will recognise that the method may also be applied to ribonucleic acid (RNA). The terms should be understood to include, as equivalents, analogs of either DNA or RNA made from nucleotide analogs and to be applicable to single stranded (such as sense or antisense) and double stranded polynucleotides. The term as used herein also encompasses cDNA, that is complementary or copy DNA produced from an RNA template, for example by the action of reverse transcriptase.

The primary polynucleotide molecules may originate in double-stranded DNA (dsDNA) form (e.g. genomic DNA fragments, PCR and amplification products and the like) or may have originated in single-stranded form, as DNA or RNA, and been converted to dsDNA form. By way of example, mRNA molecules may be copied into double-stranded cDNAs suitable for use in the method of the invention using standard techniques well known in the art. The precise sequence of the primary polynucleotide molecules is generally not material to the invention, and may be known or unknown.

In a particular embodiment, the primary polynucleotide molecules are RNA molecules. In an aspect of this embodiment, RNA isolated from specific samples is first converted to double-stranded DNA using techniques known in the art. In accordance with the present method, the double-stranded DNA is then tagged or indexed with a sample specific tag. Different preparations of such double-stranded DNA comprising sample specific tags can be generated, in parallel, from RNA isolated from different specific samples. Subsequently, different preparations of double-stranded DNA comprising different sample specific

tags can be mixed, sequenced *en masse*, and the identity of each sequenced fragment determined with respect to the sample from which it was isolated/derived by virtue of the presence of a sample specific tag.

In a particular embodiment, the primary polynucleotide molecules are DNA molecules. More particularly, the primary polynucleotide molecules represent the entire genetic complement of an organism, and are genomic DNA molecules which include both intron and exon sequences (coding sequence), as well as non-coding regulatory sequences such as promoter and enhancer sequences. Although it could be envisaged that particular sub-sets of polynucleotide sequences or genomic DNA could also be used, such as, for example, particular chromosomes. Yet more particularly, the sequence of the primary polynucleotide molecules is not known. Still yet more particularly, the primary polynucleotide molecules are human genomic DNA molecules. The DNA target molecules may be treated chemically or enzymatically either prior to, or subsequent to any random fragmentation processes, and prior to or subsequent to the ligation of the adaptor sequences.

Methods of the invention are especially useful for the parallel sequencing of a large number of relative small samples in the same experiment. In a sequencing platform with a high degree of parallelism, it may be possible to analyse 50-100 million reads in the same experiment, each to a length of 50 bases or more. Thus it may easily be possible to determine more than 5 billion base pairs of sequence from a single experiment. For a genome of 100,000-1 million base pairs, this represents a much larger coverage of each base pair than is actually needed for accurate sequencing. This feature is even more pronounced for viral or mitochondrial

genomes or collections of PCR fragments that may be 10,000 base pairs or less. For a 10 kB sample, sequenced in a run of 5 billion base pairs, each base pair will appear on average 500,000 times. For accurate sequencing, it may be desirable for each base pair to appear only 20 times, and thus it will be possible to sequence 25,000 10kB fragments in the sequencing run. The use of 8 base pair tags gives a possibility of 4^8 (65536) different samples in the same experiment. For the sake of clarity, sequencing of the tags can be performed in a separate read from the sequencing of the target fragments, so the read length or accuracy of the read from the sample is not lowered by the need to sequence the tags.

For genome sizes of 1 mB, at a depth of 20X, it is possible to sequence 250 samples in a 5 billion base pair experiment. The 250 samples can be coded by as few as a four base tag, but extra bases may be added to ensure accurate analysis. Thus, the primary polynucleotide molecules may be a genome of any organism, but the methods of the present invention may be used to particular advantage when analysing smaller genomes, such as those of bacteria, viruses, mitochondria and the like, which have genome sizes in the regions of a few million base pairs or lower.

Random fragmentation refers to the fragmentation of a polynucleotide molecule in a non-ordered fashion by enzymatic, chemical or mechanical means. Such fragmentation methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition). For the sake of clarity, generating smaller fragments of a larger piece of nucleic acid via specific PCR amplification of such smaller fragments is not equivalent to fragmenting the larger piece of nucleic acid

because the larger piece of nucleic acid sequence remains in intact (i.e., is not fragmented by the PCR amplification). Moreover, random fragmentation is designed to produce fragments irrespective of the sequence identity or position of nucleotides comprising and/or surrounding the break. More particularly, the random fragmentation is by mechanical means such as nebulisation or sonication to produce fragments of about 50 base pairs in length to about 1500 base pairs in length, still more particularly 50-700 base pairs in length, yet more particularly 50-400 base pairs in length. Most particularly, the method is used to generate smaller fragments of from 50-150 base pairs in length.

Fragmentation of polynucleotide molecules by mechanical means (nebulization, sonication and Hydroshear for example) results in fragments with a heterogeneous mix of blunt and 3'- and 5'-overhanging ends. It is therefore desirable to repair the fragment ends using methods or kits (such as the Lucigen DNA terminator End Repair Kit) known in the art to generate ends that are optimal for insertion, for example, into blunt sites of cloning vectors. In a particular embodiment, the fragment ends of the population of nucleic acids are blunt ended. More particularly, the fragment ends are blunt ended and phosphorylated. The phosphate moiety can be introduced via enzymatic treatment, for example, using polynucleotide kinase.

In a particular embodiment, the target polynucleotide sequences are prepared with single overhanging nucleotides by, for example, activity of certain types of DNA polymerase such as Taq polymerase or Klenow exo minus polymerase which has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example, deoxyadenosine (A) to the 3′ ends of, for example, PCR

products. Such enzymes can be utilised to add a single nucleotide 'A' to the blunt ended 3' terminus of each strand of the target polynucleotide duplexes. Thus, an 'A' could be added to the 3' terminus of each end repaired duplex strand of the target polynucleotide duplex by reaction with Taq or Klenow exo minus polymerase, whilst the adaptor polynucleotide construct could be a T-construct with a compatible 'T' overhang present on the 3' terminus of each duplex region of the adaptor construct. This end modification also prevents self-ligation of both vector and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

The method of the invention utilises nucleic acid sequence tags as markers characteristic of the source of particular target molecules on the array. A nucleic acid sequence tag characteristic of the source is attached to each of the target molecules in each isolated sample before the sample is immobilised for sequencing. The tag is not itself formed by part of the target nucleic acid molecule, but becomes part of the template for amplification. Generally the tag will be a synthetic sequence of nucleotides which is added to the target as part of the template preparation step. Accordingly, a sample specific tag is a nucleic acid sequence tag which is attached to each of the target molecules of a particular sample, the presence of which is indicative of or is used to identify the sample or source from which the target molecules were isolated.

Preferably the nucleic acid sequence tag may be up to 20 nucleotides in length, more preferably 1-10 nucleotides, and most preferably 4-6 nucleotides in length. A four nucleotide tag gives a possibility of multiplexing 256

samples on the same array, a six base tag enables 4096 samples to be processed on the same array.

In the first embodiment of the invention, the tag sequences are introduced in a two step process involving ligation of an adaptor that is common to all samples and amplification with a tagged amplification primer. See Figures 1 and 6. In the second embodiment the tags are on the adaptor, and the amplification can be performed with universal amplification primers. See Figure 2. In the third embodiment, the tags are on the adaptors, but at the terminus of the duplex region such that the tag and target sequences are contiguous. See Figure 3. The third embodiment allows for the sequencing to be performed in a single read, but requires both strands of the adaptor to be modified and therefore requires more oligonucleotides as two unique adaptor strands are needed for each sample.

In each embodiment, the target nucleic acids of each specific fragmented sample are treated by first ligating identical adaptor polynucleotide molecules ('mismatched adaptors', the general features of which are defined below, and further described in copending application US 20070128624, the contents of which are incorporated herein by reference in their entirety) to the 5' and 3' ends of the target polynucleotide duplexes (which may be of known, partially known or unknown sequence) to form adaptor-target constructs and then carrying out an initial primer extension reaction in which extension products complementary to both strands of each individual adaptor-target construct are formed. The resulting primer extension products, and optionally amplified copies thereof, collectively provide a library of template polynucleotides. The term library refers to the collection of target fragments containing known

common sequences at their 3' and 5' ends, and may also be referred to as a 3' and 5' modified library.

Depending on the embodiment of the invention, the adaptors may be universal for all samples, or one or both strands of the duplexes may carry the tag sequence to code or track the identity of the samples.

The adaptor polynucleotides used in the method of the invention are referred to herein as 'mismatched' adaptors because, as will be explained in detail herein, it is essential that the adaptors include a region of sequence mismatch, i.e., they must not be formed by annealing of fully complementary polynucleotide strands.

Mismatched adaptors for use in the invention are formed by annealing of two partially complementary polynucleotide strands so as to provide, when the two strands are annealed, at least one double-stranded region and at least one unmatched single-stranded region.

The 'double-stranded region' of the adaptor is a short double-stranded region, typically comprising 5 or more consecutive base pairs, formed by annealing of the two partially complementary polynucleotide strands. This term refers to a double-stranded region of nucleic acid in which the two strands are annealed and does not imply any particular structural conformation.

Generally it is advantageous for the double-stranded region to be as short as possible without loss of function. In this context, 'function' refers to the ability of the double-stranded region to form a stable duplex under standard reaction conditions for an enzyme-catalysed nucleic acid ligation reaction, which will be well known to the skilled reader (e.g. incubation at a temperature in the range of 4°C to 25°C in a ligation buffer appropriate for

the enzyme), such that the two strands forming the adaptor remain partially annealed during ligation of the adaptor to a target molecule. It is not absolutely necessary for the double-stranded region to be stable under the conditions typically used in the annealing steps of primer extension or PCR reactions.

Since identical adaptors are ligated to both ends of each target molecule, the target sequence in each adaptor-target construct will be flanked by complementary sequences derived from the double-stranded region of the adaptors. The longer the double-stranded region, and hence the complementary sequences derived therefrom in the adaptor-target constructs, the greater the possibility that the adaptor-target construct is able to fold back and base-pair to itself in these regions of internal self-complementarity under the annealing conditions used in primer extension and/or PCR. It is, therefore, generally preferred for the double-stranded region to be 20 or less, 15 or less, or 10 or less base pairs in length in order to reduce this effect. The stability of the double-stranded region may be increased, and hence its length potentially reduced, by the inclusion of non-natural nucleotides which exhibit stronger base-pairing than standard Watson-Crick base pairs.

It is preferred, but not absolutely essential, for the two strands of the adaptor to be 100% complementary in the double-stranded region. It will be appreciated that one or more nucleotide mismatches may be tolerated within the double-stranded region, provided that the two strands are capable of forming a stable duplex under standard ligation conditions.

Adaptors for use in the invention will generally include a double-stranded region forming the 'ligatable' end

of the adaptor, i.e. the end that is joined to a target polynucleotide in the ligation reaction. The ligatable end of the adaptor may be blunt or, in other embodiments, short 5' or 3' overhangs of one or more nucleotides may be present to facilitate/promote ligation. The 5' terminal nucleotide at the ligatable end of the adaptor is phosphorylated to enable phosphodiester linkage to a 3' hydroxyl group on the target polynucleotide.

The term 'unmatched region' refers to a region of the adaptor wherein the sequences of the two polynucleotide strands forming the adaptor exhibit a degree of non-complementarity such that the two strands are not capable of fully annealing to each other under standard annealing conditions for a primer extension or PCR reaction. The unmatched region(s) may exhibit some degree of annealing under standard reaction conditions for an enzyme-catalysed ligation reaction, provided that the two strands revert to single stranded form under annealing conditions in an amplification reaction.

It is to be understood that the 'unmatched region' is provided by different portions of the same two polynucleotide strands which form the double-stranded region(s). Mismatches in the adaptor construct can take the form of one strand being longer than the other, such that there is a single stranded region on one of the strands, or a sequence selected such that the two strands do not hybridise, and thus form a single stranded region on both strands. The mismatches may also take the form of 'bubbles', wherein both ends of the adaptor construct(s) are capable of hybridising to each other and forming a duplex, but the central region is not. The portion of the strand(s) forming the unmatched region are not annealed under conditions in

which other portions of the same two strands are annealed to form one or more double-stranded regions. For avoidance of doubt it is to be understood that a single-stranded or single base overhang at the 3' end of a polynucleotide duplex that subsequently undergoes ligation to the target sequences does not constitute an 'unmatched region' in the context of this invention.

The lower limit on the length of the unmatched region will typically be determined by function, for example, the need to provide a suitable sequence for binding of a primer for primer extension, PCR and/or sequencing. Theoretically there is no upper limit on the length of the unmatched region, except that in general it is advantageous to minimise the overall length of the adaptor, for example, in order to facilitate separation of unbound adaptors from adaptor-target constructs following the ligation step. Therefore, it is generally preferred that the unmatched region should be less than 50, or less than 40, or less than 30, or less than 25 consecutive nucleotides in length.

The precise nucleotide sequence of the adaptors is generally not material to the invention and may be selected by the user such that the desired sequence elements are ultimately included in the common sequences of the library of templates derived from the adaptors to, for example, provide binding sites for particular sets of universal amplification primers and/or sequencing primers. Additional sequence elements may be included, for example, to provide binding sites for sequencing primers which will ultimately be used in sequencing of template molecules in the library, or products derived from amplification of the template library, for example on a solid support.

Although the precise nucleotide sequence of the adaptor is generally non-limiting to the invention, the sequences of the individual strands in the unmatched region should be such that neither individual strand exhibits any internal self-complementarity which could lead to self-annealing, formation of hairpin structures, etc. under standard annealing conditions. Self-annealing of a strand in the unmatched region is to be avoided as it may prevent or reduce specific binding of an amplification primer to this strand.

The mismatched adaptors are preferably formed from two strands of DNA, but may include mixtures of natural and non-natural nucleotides (e.g. one or more ribonucleotides) linked by a mixture of phosphodiester and non-phosphodiester backbone linkages. Other non-nucleotide modifications may be included such as, for example, biotin moieties, blocking groups and capture moieties for attachment to a solid surface, as discussed in further detail below.

The adaptor constructs may contain exonuclease resistant modifications such as phosphorothioate linkages. Such modifications reduce the number of adaptor-dimers present in the library, since the two adaptors can not undergo ligation without removal of their non complementary overhangs. The adaptors can be treated with an exonuclease enzyme prior to the ligation reaction with the target, to ensure that the overhanging ends of the strands can not be removed during the ligation process. Treatment of the adaptors in this manner reduces the formation of the adaptor-dimers at the ligation step.

Ligation methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition). Such methods utilise

ligase enzymes such as DNA ligase to effect or catalyse joining of the ends of the two polynucleotide strands of, in this case, the adaptor duplex construct and the target polynucleotide duplexes, such that covalent linkages are formed. The adaptor duplex construct may contain a 5'-phosphate moiety in order to facilitate ligation to the target 3'-OH. The target contains a 5'-phosphate moiety, either residual from the shearing process, or added using an enzymatic treatment step, and has been end repaired, and optionally extended by an overhanging base or bases, to give a 3'-OH suitable for ligation. In this context, joining means covalent linkage of polynucleotide strands which were not previously covalently linked. In a particular aspect of the invention, such joining takes place by formation of a phosphodiester linkage between the two polynucleotide strands, but other means of covalent linkage (e.g. non-phosphodiester backbone linkages) may be used.

Optionally the combined ligated polynucleotide sequences and unligated adaptor polynucleotide constructs may be purified from any components of the ligation reaction, such as enzymes, buffers, salts and the like. Suitable purification methods are known in the art and utilise standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition).

In a next step according to the invention an amplification reaction is prepared. The contents of an amplification reaction are known by one skilled in the art and include appropriate substrates (such as dNTPs), enzymes (e.g. a DNA polymerase) and buffer components required for an amplification reaction. Generally amplification reactions require at least two amplification primers, often denoted 'forward' and 'reverse' primers (primer oligonucleotides)

that are capable of annealing specifically to a part of the polynucleotide sequence to be amplified under conditions encountered in the primer annealing step of each cycle of an amplification reaction. In certain embodiments the forward and reverse primers may be identical. Thus the primer oligonucleotides must include an 'adaptor-target specific portion', being a sequence of nucleotides capable of annealing to a part of, that is, a primer-binding sequence, in the polynucleotide molecule to be amplified (or the complement thereof if the template is viewed as a single strand) during the annealing step.

Depending on the embodiment of the invention, the amplification primers may be universal for all samples, or one of the forward or reverse primers may carry the tag sequence that codes for the sample source. The amplification primers may hybridise across the tag region of the ligated adaptor, in which case unique primers will be needed for each sample nucleic acid. The amplification reaction may be performed with more than two amplification primers. In order to prevent the amplification of ligated adapter-adapter dimers, the amplification primers can be modified to contain nucleotides that hybridise across the whole of the ligated adapter and into the ligated template (or the dNTP's attached to the 3' end thereof). This first amplification primer can be modified and treated to help prevent exonucleases digestion of the strands, and thus it may be advantageous to have a first amplification primer that is universal and can amplify all samples rather than modifying and treating each of the tagged primers separately. The tagged primer can be introduced as a sample specific third primer in the amplification reaction, but does not need to be specially modified and treated to lower exonuclease

digestion. The nested PCR approach is shown in figure 6. In the case of this embodiment the third amplification primer that carries the tag needs to contain at sequence the same as at least a portion of the first amplification primer such that it can be used to amplify the duplex resulting from extension of the first amplification primer.

In the context of the present invention, the term 'polynucleotide molecule to be amplified' refers to the original or starting adaptor-target-adaptor sequence added to the amplification reaction. The 'adaptor-target specific portion' in the forward and reverse amplification primers refers to a sequence capable of annealing to the original or initial adaptor-target-adaptor present at the start of the amplification reaction and reference to the length of the 'adaptor-target specific portion' relates to the length of the sequence in the primer which anneals to the starting adaptor-target. It will be appreciated that if the primers contain any nucleotide sequence which does not anneal to the starting adaptor-target in the first amplification cycle then this sequence may be copied into the amplification products (assuming the primer does not contain a moiety which prevents read-through of the polymerase). Hence the amplified template strands produced in the first and subsequent cycles of amplification may be longer than the starting adaptor-target strands.

As the mis-matched adapters can be different lengths, the length of adapter sequence added to the 3' and 5' ends of each strand may be different. The amplification primers may also be of different length to each other, and may hybridise to different lengths of the adapter, and therefore the length added to the ends of each strand can be controlled. In the case of nested PCR, the three or more

amplification primers can be designed to be longer than the primer used to amplify the previous amplicon, so the length of the added nucleotides is fully controllable and may be hundreds of base pairs if desired. In the example shown in figure 6b, the first amplification primer adds 13 bases to the ligated adapter, and the third amplification primer adds a further 27 bases such that one end of the amplicon is 40 bases longer than the short arm of adapter-target construct. The short arm of the adapter is 20 bases in length, meaning that the prepared template comprises the genomic region plus 60 added bases at the end. In figure 6b, the second amplification primer is 25 bases longer than the long arm of adapter, which is 32 bases in length plus the additional T that hybridise across the dATP nucleoside added to the genomic sample. Thus the prepared template comprises the genomic fragment, plus the added dATP, plus 57 known bases. Thus in full, one strand of each template duplex comprises from the 5' end: 60 known bases, T, the genomic fragment, A, 57 known bases-3' end. This strand is fully complementary to a sequence: 5'-57 known bases, T, genomic fragment, A, 60 known bases-3' end. The length 57 and 6 are arbitary, and shown for the purpose of clarification, and should not be taken as limiting. The length of the added sequences may be 20-100 bases or more depending on the desired experimental design.

The forward and reverse primers may be of sufficient length to hybridise to the whole of the adaptor sequence and at least one base of the target sequence (or the nucleotide dNTP added as a 3'-overhang on the target strands). The forward and reverse primers may also contain a region that extends beyond the adaptor construct, and therefore the amplification primers may be at least 20-100 bases in

length. The forward and reverse primers may be of significantly different lengths; for example one may be 20-40 bases, whereas the other one may be 40-100 bases in length. The nucleotide sequences of the adaptor-target specific portions of the forward and reverse primers are selected to achieve specific hybridisation to the adaptor-target sequences to be amplified under the conditions of the annealing steps of the amplification reaction, whilst minimising non-specific hybridisation to any other target sequences present.

Skilled readers will appreciate that it is not strictly required for the adaptor-target specific portion to be 100% complementary, a satisfactory level of specific annealing can be achieved with less than perfectly complementary sequences. In particular, one or two mis-matches in the adaptor-target specific portion can usually be tolerated without adversely affecting specificity for the template. Therefore the term 'adaptor-target specific portion' should not be interpreted as requiring 100% complementarity with the adaptor-target. However, the requirement that the primers do not anneal non-specifically to regions of the adaptor-target other than their respective primer-binding sequences must be fulfilled.

Amplification primers are generally single stranded polynucleotide structures. They may also contain a mixture of natural and non-natural bases and also natural and non-natural backbone linkages, provided that any non-natural modifications do not preclude function as a primer - that being defined as the ability to anneal to a template polynucleotide strand during conditions of the amplification reaction and to act as an initiation point for synthesis of

a new polynucleotide strand complementary to the template strand.

Primers may additionally comprise non-nucleotide chemical modifications, for example phosphorothioates to increase exonuclease resistance, again provided such that modifications do not prevent primer function. Modifications may, for example, facilitate attachment of the primer to a solid support, for example a biotin moiety. Certain modifications may themselves improve the function of the molecule as a primer, or may provide some other useful functionality, such as providing a site for cleavage to enable the primer (or an extended polynucleotide strand derived therefrom) to be cleaved from a solid support.

In an embodiment wherein tags are attached to the adaptors, the amplification can be carried out on either the pooled or unpooled samples. In an embodiment wherein universal adaptors are used, tags must be part of the amplification primers, and therefore, each sample must be amplified independently prior to pooling. The pooled nucleic acid samples can then be processed for sequencing

Preparation of immobilised samples for sequencing

The pooled sample can then be immobilised in preparation for sequencing. Sequencing can be performed as an array of single molecules, or can be amplified prior to sequencing. The amplification can be carried out using one or more immobilised primers. The immobilised primer(s) can be a lawn on a planar surface, or on a pool of beads. The pool of beads can be isolated into an emulsion with a single bead in each 'compartment' of the emulsion. At a concentration of only one template per 'compartment', only a single template is amplified on each bead.

The term 'solid-phase amplification' as used herein refers to any nucleic acid amplification reaction carried out on or in association with a solid support such that all or a portion of the amplified products are immobilised on the solid support as they are formed. In particular, the term encompasses solid-phase polymerase chain reaction (solid-phase PCR) and solid phase isothermal amplification which are reactions analogous to standard solution phase amplification, except that one or both of the forward and reverse amplification primers is/are immobilised on the solid support. Solid phase PCR covers systems such as emulsions, wherein one primer is anchored to a bead and the other is in free solution, and colony formation in solid phase gel matrices wherein one primer is anchored to the surface, and one is in free solution.

Although the invention encompasses 'solid-phase' amplification methods in which only one amplification primer is immobilised (the other primer usually being present in free solution), it is preferred for the solid support to be provided with both the forward and the reverse primers immobilised. In practice, there will be a 'plurality' of identical forward primers and/or a 'plurality' of identical reverse primers immobilised on the solid support, since the amplification process requires an excess of primers to sustain amplification. References herein to forward and reverse primers are to be interpreted accordingly as encompassing a 'plurality' of such primers unless the context indicates otherwise.

As will be appreciated by the skilled reader, any given amplification reaction requires at least one type of forward primer and at least one type of reverse primer specific for the template to be amplified. However, in certain

embodiments the forward and reverse primers may comprise template-specific portions of identical sequence, and may have entirely identical nucleotide sequence and structure (including any non-nucleotide modifications). In other words, it is possible to carry out solid-phase amplification using only one type of primer, and such single-primer methods are encompassed within the scope of the invention. Other embodiments may use forward and reverse primers which contain identical template-specific sequences but which differ in some other structural features. For example one type of primer may contain a non-nucleotide modification which is not present in the other.

In all embodiments of the invention, primers for solid-phase amplification are preferably immobilised by single point covalent attachment to the solid support at or near the 5' end of the primer, leaving the template-specific portion of the primer free to anneal to its cognate template and the 3' hydroxyl group free for primer extension. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivatisation or functionalisation applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In a particular embodiment, the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5' end. In the case of solid-supported polyacrylamide hydrogels (as described below), this nucleophile will bind to a bromoacetamide group present in the hydrogel. A more particular means of attaching primers and templates to a solid support is via 5' phosphorothioate attachment to a hydrogel comprised of polymerised acrylamide

and *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA), as described fully in WO05065814.

Certain embodiments of the invention may make use of solid supports comprised of an inert substrate or matrix (e.g. glass slides, polymer beads, etc) which has been "functionalised", for example by application of a layer or coating of an intermediate material comprising reactive groups which permit covalent attachment to biomolecules, such as polynucleotides. Examples of such supports include, but are not limited to, polyacrylamide hydrogels supported on an inert substrate such as glass. In such embodiments, the biomolecules (e.g. polynucleotides) may be directly covalently attached to the intermediate material (e.g. the hydrogel), but the intermediate material may itself be non-covalently attached to the substrate or matrix (e.g. the glass substrate). The term "covalent attachment to a solid support" is to be interpreted accordingly as encompassing this type of arrangement.

The pooled samples may be amplified on beads wherein each bead contains a forward and reverse amplification primer. In a particular embodiment, the library of templates prepared according to the first, second or third aspects of the invention is used to prepare clustered arrays of nucleic acid colonies, analogous to those described in US20050100900, US7115400, WO 00/18957 and WO 98/44151, the contents of which are incorporated herein by reference in their entirety, by solid-phase amplification and more particularly solid phase isothermal amplification. The terms 'cluster' and 'colony' are used interchangeably herein to refer to a discrete site on a solid support comprised of a plurality of identical immobilised nucleic acid strands and a plurality of identical immobilised complementary nucleic

acid strands. The term 'clustered array' refers to an array formed from such clusters or colonies. In this context the term 'array' is not to be understood as requiring an ordered arrangement of clusters.

The term solid phase, or surface, is used to mean either a planar array wherein primers are attached to a flat surface, for example, glass, silica or plastic microscope slides or similar flow cell devices; beads, wherein either one or two primers are attached to the beads and the beads are amplified; or an array of beads on a surface after the beads have been amplified.

Clustered arrays can be prepared using either a process of thermocycling, as described in WO9844151, or a process whereby the temperature is maintained as a constant, and the cycles of extension and denaturing are performed using changes of reagents. Such isothermal amplification methods are described in patent application numbers WO0246456and US20080009420 (Isothermal methods for creating clonal single molecule arrays), which are incorporated herein by reference in their entirety. Due to the lower temperatures required in the isothermal process, this is particularly preferred.

Use in sequencing/methods of sequencing

The invention also encompasses methods of sequencing amplified nucleic acids generated by solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a pool of nucleic acid templates using solid-phase amplification as described above and carrying out a nucleic acid sequencing reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the

solid-phase amplification reaction, as well as the sequence of the tag that codes for the identity of the sample.

Sequencing can be carried out using any suitable sequencing technique, wherein nucleotides are added successively to a free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added is preferably determined after each nucleotide addition. Sequencing techniques using sequencing by ligation, wherein not every contiguous base is sequenced, and techniques such as massively parallel signature sequencing (MPSS) where bases are removed from, rather than added to the strands on the surface are also within the scope of the invention, as are techniques using detection of pyrophosphate release (pyrosequencing). Such pyrosequencing based techniques are particularly applicable to sequencing arrays of beads where the beads have been amplified in an emulsion such that a single template from the library molecule is amplified on each bead.

The initiation point for the sequencing reaction may be provided by annealing of a sequencing primer to a product of the solid-phase amplification reaction. In this connection, one or both of the adaptors added during formation of the template library may include a nucleotide sequence which permits annealing of a sequencing primer to amplified products derived by whole genome or solid-phase amplification of the template library.

Depending on the embodiment of the invention, the tag sequence and fragment sequence may be determined in a single read from a single sequencing primer, or in multiple reads from two sequencing primers. In the case of two reads from two sequencing primers, the 'tag read' and the 'fragment

read' can be performed in either order, with a suitable denaturing step to remove the annealed primer after the first sequencing read is completed. Suitable denaturing steps may include formamide, hydroxide or heat as described elsewhere.

The products of solid-phase amplification reactions wherein both forward and reverse amplification primers are covalently immobilised on the solid surface are so-called 'bridged' structures formed by annealing of pairs of immobilised polynucleotide strands and immobilised complementary strands, both strands being attached to the solid support at the 5' end. Arrays comprised of such bridged structures provide inefficient templates for nucleic acid sequencing, since hybridisation of a conventional sequencing primer to one of the immobilised strands is not favoured compared to annealing of this strand to its immobilised complementary strand under standard conditions for hybridisation.

In order to provide more suitable templates for nucleic acid sequencing, it is preferred to remove substantially all or remove or displace at least a portion of one of the immobilised strands in the 'bridged' structure in order to generate a template which is at least partially single-stranded. The portion of the template which is single-stranded will thus be available for hybridisation to a sequencing primer. The process of removing all or a portion of one immobilised strand in a 'bridged' double-stranded nucleic acid structure may be referred to herein as 'linearisation', and is described in further detail in WO07010251, the contents of which are incorporated herein by reference in their entirety.

Bridged template structures may be linearised by cleavage of one or both strands with a restriction endonuclease or by cleavage of one strand with a nicking endonuclease. Other methods of cleavage can be used as an alternative to restriction enzymes or nicking enzymes, including *inter alia* chemical cleavage (e.g. cleavage of a diol linkage with periodate), cleavage of abasic sites by cleavage with endonuclease (for example 'USER', as supplied by NEB, part number M5505S), or by exposure to heat or alkali, cleavage of ribonucleotides incorporated into amplification products otherwise comprised of deoxyribonucleotides, photochemical cleavage or cleavage of a peptide linker.

It will be appreciated that a linearization step may not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilised and the other in free solution.

Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may be subjected to denaturing conditions in order to remove the portion(s) of the cleaved strand(s) that are not attached to the solid support. Suitable denaturing conditions, for example sodium hydroxide solution, formamide solution or heat, will be apparent to the skilled reader with reference to standard molecular biology protocols (Sambrook *et al.*, 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Denaturation results in the production of a sequencing template which is partially or substantially single-stranded. A sequencing reaction may then be initiated by hybridisation of a sequencing primer to the single-stranded portion of the template.

Thus, the invention encompasses methods wherein the nucleic acid sequencing reaction comprises hybridising a sequencing primer to a single-stranded region of a linearised amplification product, sequentially incorporating one or more nucleotides into a polynucleotide strand complementary to the region of amplified template strand to be sequenced, identifying the base present in one or more of the incorporated nucleotide(s) and thereby determining the sequence of a region of the template strand.

One preferred sequencing method which can be used in accordance with the invention relies on the use of modified nucleotides having removable 3′ blocks, for example as described in WO04018497 and US7057026, the contents of which are incorporated herein by reference in their entirety. Once the modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase can not add further nucleotides. Once the nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides, it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has a different label attached thereto, known to correspond to the particular base, to facilitate discrimination between the bases added during each incorporation step. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides separately.

The modified nucleotides may carry a label to facilitate their detection. A fluorescent label, for example, may be used for detection of modified nucleotides. Each nucleotide type may thus carry a different fluorescent label, for example, as described in US Provisional Application No. 60/801,270 (Novel dyes and the use of their labelled conjugates), published as WO07135368, the contents of which are incorporated herein by reference in their entirety. The detectable label need not, however, be a fluorescent label. Any label can be used which allows the detection of an incorporated nucleotide.

One method for detecting fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of illumination. The fluorescence from the label on the nucleotide may be detected by a CCD camera or other suitable detection means. Suitable instrumentation for recording images of clustered arrays is described in US Provisional Application No. 60/788,248 (Systems and devices for sequence by synthesis analysis), published as WO07123744, the contents of which are incorporated herein by reference in their entirety.

The invention is not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used. Suitable alternative techniques include, for example, Pyrosequencing™, FISSEQ (fluorescent in situ sequencing), MPSS and sequencing by ligation-based methods, for example as described is US6306597.

The nucleic acid sample may be further analysed to obtain a second read from the opposite end of the fragment.

Methodology for sequencing both ends of a cluster are described in co-pending applications WO07010252 and PCTGB2007/003798, the contents of which are incorporated by reference herein in their entirety. In one example, the series of steps may be performed as follows; generate clusters, linearise, hybridise first sequencing primer and obtain first sequencing read. The first sequencing primer can be removed, a second primer hybridised and the tag sequenced. The nucleic acid strand may then be 'inverted' on the surface by synthesising a complementary copy from the remaining immobilised primers used in cluster amplification. This process of strand resynthesis regenerates the double stranded cluster. The original template strand can be removed, to linearise the resynthesised strand that can then be annealed to a sequencing primer and sequenced in a third sequencing run.

In the cases where strand resynthesis is employed, both strands must be immobilised to the surface in a way that allows subsequent release of a portion of the immobilised strand. This can be achieved through a number of mechanisms as described in WO07010251, the contents of which are incorporated herein by reference in their entirety. For example, one primer can contain a uracil nucleotide, which means that the strand can be cleaved at the uracil base using the enzymes uracil glycosylase (UDG) which removes the nucleoside base, and endonuclease VIII that excises the abasic nucleotide. This enzyme combination is available as USER™ from New England Biolabs (NEB part number M5505). The second primer may comprise an 8-oxoguanine nucleotide, which is then cleavable by the enzyme FPG (NEB part number M0240). This design of primers give complete control of which primer is cleaved at which point in the process, and also where in the cluster the cleavage occurs. The primers may also be chemically modified, for example with a disulfide or diol

modification that allows chemical cleave at specific locations.

Kits

The invention also relates to kits for use in preparing libraries of tagged polynucleotides using the method of the first, second or third aspects of the invention.

Embodiments of the kit comprise at least a supply of a mismatched adaptor as defined herein, plus a supply of at least one amplification primer which is capable of annealing to the mismatched adaptor and priming synthesis of an extension product, which extension product would include any target sequence ligated to the adaptor when the adaptor is in use.

Particular features of the 'mismatch' adaptors for inclusion in the kit are as described elsewhere herein in relation to other aspects of the invention. The structure and properties of amplification primers are well known to those skilled in the art. Suitable primers of appropriate nucleotide sequence for use with the adaptors included in the kit can be readily prepared using standard automated nucleic acid synthesis equipment and reagents in routine use in the art. The kit may include a supply of one single type of primer or separate supplies (or even a mixture) of two different primers, for example a pair of amplification primers suitable for PCR or isothermal amplification of templates modified with the mismatched adaptor in solution phase and/or on a suitable solid support (i.e. solid-phase amplification). The kit may comprise a mismatch double stranded adapter for ligation to a sample of interest, plus

at least two different amplification primers that carry a different tag sequence, where the tag sequence does not hybridise to the adapter. This kit can be used to amplify at least two different samples where each sample is amplified using a single tagged primer, and then pooled after the individual amplification reactions.

In one embodiment the kit may include supplies of different primer-pairs for use in solution phase and solid phase PCR and more particularly isothermal amplification. In this context the 'different' primer-pairs may be of substantially identical nucleotide sequence but differ with respect to some other feature or modification, such as for example surface-capture moieties, etc. In other embodiments the kit may include a supply of primers for use in an initial primer extension reaction and a different primer-pair (or pairs) for solution and/or solid phase amplification.

Adaptors and/or primers may be supplied in the kits ready for use, or more preferably as concentrates requiring dilution before use, or even in a lyophilised or dried form requiring reconstitution prior to use. If required, the kits may further include a supply of a suitable diluent for dilution or reconstitution of the primers. Optionally, the kits may further comprise supplies of reagents, buffers, enzymes, dNTPs, etc., for use in carrying out PCR or isothermal amplification. Suitable (but non-limiting) examples of such reagents are as described in the Materials and Methods sections of the accompanying Examples. Further components which may optionally be supplied in the kit include 'universal' sequencing primers suitable for sequencing templates prepared using the mismatched adaptors and primers.

The invention will be further understood with reference to the following non-limiting experimental examples:

Examples

The following are examples of general techniques which may be applied in carrying out the method of the invention.

Example 1: Preparation and conventional sequencing of tagged libraries:

The following experimental details describe the complete exposition of one embodiment of the invention as described above.

Two libraries were made, one using purified human BAC DNA (140k human chromosome 6 insert cloned into a pTARBAC vector) and the other purified PhiX174 RF1 DNA (NEB, N3021L). Each library was prepared separately. The DNA was first prepared for ligation to forked adaptors by: fragmentation of the DNA by nebulisation, end repair of the DNA ends to render them blunt-ended and phosphorylated, then the addition of a single 'A' nucleotide onto the 3' ends of the DNA fragments. The ligation reaction was performed with the prepared fragmented DNA and adaptors pre-formed by annealing 'Oligo A' and 'Oligo B' (sequences given below). The same adaptors were used for each library. The product of the reaction was isolated/purified from unligated adaptor by gel electrophoresis. Finally, the product of the ligation reaction was subjected to cycles of PCR to selectively amplify ligated product that contained genomic DNA with adaptor at both ends of the fragments. During the PCR a unique tag was added to each library using a unique PCR

primer, so that the BAC library was indexed with the unique sequence tag 'ATC' and the PhiX library was indexed with the unique sequence tag 'CGA'. See Figure 1 for general schematic of protocol.

Materials and Methods

Step 1) Nebulization

Materials:

* 0.5 ug/ul Human BAC DNA (140k human chromosome 6 insert cloned into a pTARBAC vector)
* 1ug/ul PhiX174 RF1 DNA (NEB, N3021L)
* Nebulization Buffer (53.1 ml glycerol, 42.1 ml water, 3.7 ml 1 M Tris HCl pH7.5, 1.1 ml 0.5 M EDTA)
* TE
* Nebulizers (Invitrogen, K7025-05)
* PCR purification kit columns (Qiagen, 28104)

Procedure:

Mixed 5 µl (5 µg) of PhiX174 DNA with 45 µl of TE and 700 µl l of nebulization buffer. Mixed 10 µl (5 µg) of BAC DNA with 40 µl of TE and 700 µl of nebulization buffer. Chilled DNA solutions were each fragmented in a nebulizer on ice for 6 minutes under 32 pounds per square inch (psi) of pressure. The recovered volumes were each purified with a Qiagen PCR purification kit column and eluted in 30 µl of EB.

Step 2) End-repair

Materials:

* Nebulized DNA (from Step 1)
* Water

- T4 DNA ligase buffer with 10mM ATP (10x) (NEB, B0202S)
- dNTPs mix (10mM each) (NEB, N0447S)
- T4 DNA Polymerase (3U/ul) (NEB, M0203L)
- *E. coli* DNA Pol I large fragment (Klenow) (5U/ul) (NEB, M0210S)
- T4 polynucleotide kinase (10U/ul) (NEB, M0201L)
- PCR purification kit columns (Qiagen, 28104)

Procedure:

End repair mix was assembled as follows:

Nebulized DNA	30µl
Water	45µl
T4 DNA ligase buffer with 10mM ATP	10µl
dNTPs	4µl
T4 DNA pol	5µl
Klenow DNA pol	1ul
T4 PNK	5ul
	100µl total

The reaction was incubated for 30 minutes at room temperature. The DNA was purified on a Qiagen column, eluting in 30µl EB.

Step 3) A- tailing Reaction

Materials:

- End repaired DNA (from Step 2)
- Water
- NEB buffer 2 (10x) (NEB, B7002S)
- dATP (1mM) (Amersham-Pharmacia, 272050)
- Klenow fragment (3' to 5' exo minus) (5U/ul) (NEB, M0212B)

* Hot block or PCR machine
* MinElute PCR purification kit column (Qiagen, 28004)

Procedure:

The following reaction mix was assembled:

End repaired DNA	30µl
Water	2ul
NEB buffer 2	5µl
dATP	10µl
Klenow fragment (3' to 5' exo minus)	3µl
	50µl total

The reaction was incubated for 30 minutes at 37°C, then the DNA purified on a Qiagen MinElute column, eluting in 10µl EB.

Step 4) Annealed Adaptors

Materials:

* Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCxT (x = phosphorothioate bond) (SEQ ID No. 1)
* Oligo B: 5'Phosphate-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG (SEQ ID No. 2)
* 50mM Tris/50mM NaCl pH7.0
* PCR machine

Procedure:

The oligos were mixed together to a final concentration of 15uM each, in 10mM Tris/10mM NaCl pH 7.0. The adaptor strands were annealed in a PCR machine programmed as follows: Ramp at 0.5 °C/sec to 97.5 °C; Hold at 97.5 °C for 150 sec; then a step of 97.5 °C for 2 sec with a temperature drop of 0.1 °C/cycle for 775 cycles.

In order to ensure a sequencing primer is attached to both ends of the adapter for the paired end method, the sequence of oligonucleotide B was changed to:

5' Phosphate-GATCGGAAGAGCACACGTCT 3' (SEQ ID No. 3)

Step 5) Ligation

Materials:

- A-tailed genomic DNA (from Step 3)
- Quick ligase buffer (2x) (NEB, B2200S)
- Annealed adaptor (15uM) (from 4.)
- Quick Ligase (1U/ul) (NEB, M2200L)
- PCR purification kit columns (Qiagen, 28104)

Procedure:

Reaction mix was assembled as follows:

A-tailed genomic DNA	10µl
Quick ligase buffer	25µl
Annealed adaptor	10µl
Quick Ligase	5µl
	50µl total

The reaction was incubated for 15 minutes at room temperature, then the DNA purified on a Qiagen column, eluting in 30µl EB.

Step 6) Gel purification

Materials:

- Ligation reaction (from Step 5)
- Agarose (Biorad, 161-3107)
- TAE (50x)

* Distilled water
* Ethidium bromide (Sigma, E1510)
* Loading buffer (4x) (50 mM Tris pH8, 40 mM EDTA, 40% w/v sucrose)
* Low molecular weight ladder (NEB, N3233L)
* Gel trays and tank. Electrophoresis unit
* Dark reader transilluminator (Clare Chemical Research, D195M)
* Gel extraction kit columns (Qiagen, 28704)

Procedure:

The entire sample from the purified ligation reaction was loaded into one lane of a 2% agarose gel containing ethidium bromide and run at 120V for 60 minutes. The gel was then viewed on a 'White-light' box and fragments from 120bp to 170bp excised and purified with a gel extraction column, eluting in 30µl elution buffer (EB).

Step 7) Exonuclease I Treatment of PCR Primers

Materials:

* Exonuclease I (E. coli) (20 U/ul) (NEB, M0293S)
* Exonuclease I Reaction Buffer (10x) (NEB, M0293S)
* Water
* DNA Primers with a phosphorothioate at the n-1 position
* P6 Bio-Rad columns (Bio-Rad, 732-6221)

Procedure:

DNA Primers with a phosphorothioate at the n-1 position (5 x 85 µl of each Primer (approx 25 µM)) were aliquoted into Eppendorf tubes. 10 µl of 10 X Exonuclease I Reaction Buffer and 5 µl of Exonuclease I was added to each tube.

Each Eppendorf tube was placed in a rack and stored in an oven set at 37 °C for 16 hours. After 16 hr, the tubes were placed on a hotblock set at 80 °C for 2 minutes. Subsequently, the solutions from the Eppendorfs were passed through P6 Bio-Rad columns and spun in a centrifuge at 2000 rpm for 2 minutes. An extra 20 µl of H_20 was added to the columns and the columns respun. The filtered solutions were placed into a SpeedVac® and evaporated until each was at 20 µl, and the fractions combined. The pooled fractions were injected into a reverse phase HPLC system, and the main peak was collected. The collected fractions were evaporated to dryness in a SpeedVac®, 50 µl of water was added and the fraction was subjected again to evaporation to dryness. The resulting pellets were dissolved in 50 µl of water, pooled and a UV measurement taken to determine the concentration of the oligonucleotide.

Step 8) PCR

Materials:

- Gel purified DNA (from Step 6)
- Water
- Phusion master mix (2x) (NEB, F-531L)
- Exonuclease treated universal PCR primer (25uM): 5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA TCxT (SEQ ID No. 4) 3', where x= phosphorothioate bond (from Step 7)
- Exonuclease treated indexing PCR primer 1 (25uM): 5' CAAGCAGAAGACGGCATACGA**GAT**CGGTCTCGGCATTCCTGCTGAACCGCTCTTC CGATCxT (SEQ ID No. 5), where x= phosphorothioate bond (from Step 7)
- Exonuclease treated indexing PCR primer 2 (25uM): 5' CAAGCAGAAGACGGCATACGA**TCG**CGGTCTCGGCATTCCTGCTGAACCGCTCTTC

CGATCxT (SEQ ID No. 6) , where x= phosphorothioate bond (from Step 7)

* PCR machine
* PCR purification kit columns (Qiagen, #28104)

Procedure:

The PCR reaction was prepared as follows:

BAC Library	
Gel purified BAC DNA	1µl
Phusion mastermix	25µl
Universal PCR primer	1µl
Indexing PCR primer 1	1µl
Water	22µl
	50µl total

PhiX174 Library	
Gel purified PhiX174 DNA	1µl
Phusion mastermix	25µl
Universal PCR primer	1µl
Indexing PCR primer 2	1µl
Water	22µl
	50µl total

Thermocycling was carried out in a PCR machine under the following conditions:

* 30 secs @ 98°C
* [10 sec@ 98°C, 30 sec @ 65°C, 30 sec @ 72°C] 18 cycles
* 5 minutes @ 72°C
* Hold @ 4°C

PCR products were purified on a Qiagen column, eluting in 30µl EB. The resulting DNA libraries were ready for

mixing and amplification on a surface amplification platform.

In the case of paired end runs, the human and Phi-X sample were split into 6 portions, and each portion was amplified with the two universal primers plus one of the 12 primers shown below.

Universal LPX primer 5'
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCxT (SEQ ID No. 7) 3'
Lpx primer 5'
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT (SEQ ID No. 8)
Taq1 primer (CGTGAT) 5'
CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC (SEQ ID No. 9) 3'
Taq2 primer (ACATCG) 5'
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC (SEQ ID No. 10) 3'
Taq3 primer (GCCTAA) 5'
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTC (SEQ ID No. 11) 3'
Taq4 primer (TGGTCA) 5'
CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC (SEQ ID No. 12) 3'
Taq5 primer (CACTGT) 5'
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC (SEQ ID No. 13) 3'
Taq6 primer (ATTGGC) 5'
CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC (SEQ ID No. 14) 3'

Tag7 primer (GATCTG) 5′
CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTC (SEQ ID No. 15)
3′
Tag8 primer (TCAAGT) 5′
CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC (SEQ ID No. 16)
3′
Tag9 primer (CTGATC) 5′
CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTC (SEQ ID No. 17)
3′
Tag10 primer (AAGCTA) 5′
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTC (SEQ ID No. 18)
3′
Tag11 primer (GTAGCC) 5′
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC (SEQ ID No. 19)
3′
Tag12 primer (TACAAG) 5′
CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTC (SEQ ID No. 20)
3′

The two universal primers contain a phosphorothioate group (x) and were exonucleases treated and purified prior to amplification. The amplification conditions were as described above, and the 6 tagged human samples, and 6 tagged Phi-X sample were pooled together for sequencing on clusters.

Validation of Libraries by Conventional Sanger Sequencing

Four (4) µl of the libraries were cloned into a plasmid vector (Zero Blunt TOPO PCR cloning kit, Invitrogen #K2800-20) and plated out on agar, according to the manufacturer's instructions. Colonies were picked, mini-prepped and the

cloned fragments sequenced by conventional Sanger sequencing.

16 clones from BAC Library ([redacted] tag)

1 (204bp) Insert: E.coli 85bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTACTGATTTCATTGCAGCCAAAGGCAAACTTTGGCTGCATCGTTTTA
CAGTCGCCATAAGCCTTTCCTCTGTTAAACCGCCTTCTGAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCTGCTTG
(SEQ ID No. 21)

2 (214bp) Insert: BAC 95bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTATCAATATTGTGAAAATGACCATACTGCCAAAAAAAAACTACAAATT
CAATGCAATTTTCATCAAAATACCATCATCATTCTTCACAATATTGATAGATC
GGAAGAGCGGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCT
GCTTG (SEQ ID No. 22)

3 (215bp) Insert: BAC 96bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTCACTCCTGGCAGAGGGACGTGTGACTAGCCATGGGCCCCTAG
GTCTCCAGTTCCTGGGTAGCTTGTATTTTTGAACATCTCCTGTATATTAGTTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 23)

4 (147bp) Insert: BAC 28bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAGTGTAGTTGAGATCTGCCTTAGCAGCAAGATCGGAAGAGCGGTT
CAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 24)

5 (183bp) Insert: BAC 64bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAACACATTTCAAAGTTTGGGGCCCTCCTCCTCCCCAAAAAACAAAC
CACAAAAAACAAACAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
ACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 25)

6 (170bp) Insert: BAC 59bp

GGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTG
AATGCCTTTTATAGCATTTAATTTTTCCTAAGTATAATTACCAAATAAAAATTG
TATAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATG
CCGTCTTCTGCTTG (SEQ ID No. 26)

7 (180bp) Insert: BAC 61bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGGGCCCGGGAGGAGTTTGCCGGGGAGGAGTGGGTTTGGAATCG
GGGTTAAAGGAAAGAGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
ACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 27)

8 (190bp) Insert: BAC 73bp

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTAAGATCTATTTCAAATGGACTGTAGATCTAAGTATAAAAGGTAAGAGA
ATAATTATTCTAGAAAGTAAATGTAAGATCGGAAGAGCGGTTCAGCAGGAAT
GCCGAGACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 28)

9 (192bp) Insert: BAC 74bp

AATGATACGGCGACCACCAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTGGGAGGCCAAGGTGGGTGGATCACCTGAGATCAGGAGTTCGAGAC
CAGCTGGCCAACATGATGAAACTCTGTCTAGATCGGAAGAGCGGTTCAGCA
GGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No.
29)

IDT Exhibit 1018.0694

10 (185bp) Insert: BAC 66bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGACCATTGTAACCATTAATGTAGACTGCAATGATATGCACTATTTA
CAACCTTTTTTAAGACTCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
GACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No. 30)

11 (199bp) Insert: BAC 80bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTTGAAGAGCTGGCAGTAGAAGATAAACAGGCTGGGGAAGAAGA
GAAAGTGCTCAAGGAGAAGGAGCAGCAGCAGCAGC
AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 31)

12 (212bp) Insert: BAC 93bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAGTATTCAACAAGTCTGTCTTTTCCAAGTGTCTTTAAAGACCAGAAA
TACCTGTTTTTAACACACAGGGTTGCAAAATTCAGAGGAGATTGGCAGATCG
GAAGAGCGGTTCAGCAGGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCTG
CTTG (SEQ ID No. 32)

13 (247bp) Insert: E.coli 128bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCTTGAGATGAGTGATGACGGCGCGCTGGAAGTTGCTCGTCGCGCT
CGCGGTACGCCGCGCATTGCCAACCGTCTGCTGCGTCGAGTGCGTGATTTC
GCCGAAGTGAAGCACGATGGCACCATCTCAAGAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCG[redacted]TCGTATGCCGTCTTCTGCTTG (SEQ ID No.
33)

14 (202bp) Insert: BAC 83bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGGGTTGGTGGAACCCAGATGCCTCCCAGGATTGGTGGGCCCTG

TGGCACTTGTACCTGCTGTTGCTGTTGCTGCTGCTGCTGAGATCGGAAGAG
CGGTTCAGCAGGAATGCCGAGACCGATCGTCGTATGCCGTCTTCTGCTTG (SE
Q ID No. 34)

15 (166bp) Insert: BAC 47bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTATGATAAGGAGCAGGTTTACAGATCATAAGTGCAAAAGCGGGCGA
GAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCGTCGTATGC
CGTCTTCTGCTTG (SEQ ID No. 35)

16 (147bp) Insert: BAC 31bp
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCTGATACTGTTGTAACCACCCAATTGGTTCAAGATCGGAAGAGCGGTT
CAGCAGGAATGCCGAGACCGATCGTCGTATGCCGTCTTCTGCTTG (SEQ ID
No. 36)

16 clones from PhiX174 Library (CGA tag)

1 (183bp) Insert: PhiX 67bp
AATGATACGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTA
ACAAAAAGTCAGATATGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGA
CCGCGATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 37)

2 (174bp) Insert: PhiX 58bp
GATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCTCGACCACTCGCGATTCAATCATGACTTCGTGATAAAAGATTGAGTGTG
AGGTTATAACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCGAT
CGTATGCCGTCTTCTGCTTG (SEQ ID No. 38)

3 (179bp) Insert: PhiX 62bp

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCTTCTATGAAGGATGTTTTCCGTTCTGGTGATTCGTCTAAGAAGTTTAAG
ATTGCTGAGGGTCAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC
GCCATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 39)

4 (169bp) Insert: PhiX 50bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGTTGCGCCGCCAAAACGTCGGCTACAGTAACTTTTCCCAGCCTC
AATCTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCCATCGTAT
GCCGTCTTCTGCTTG (SEQ ID No. 40)

5 (166bp) Insert: PhiX 50bp
TGATACGGCGACCACCGAGATCTACATCTTTCCCTACACGACGCTCTTCCGA
TCTAACAACTGAACGGACTGGAAACACTGGTCATAATCATGGTGGCGAATAA
GAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCCATCGTATGCC
GTCTTCTGCTTG (SEQ ID No. 41)

6 (164bp) Insert: PhiX 45bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAAATACGTTAACAAAAAGTCAGATATGGACCTTGCTGCTAAAGGTA
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCCATCGTATGCCGT
CTTCTGCTTG (SEQ ID No. 42)

7 (194bp) Insert: PhiX 75bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGTCAATATAACCAGTAGTGTTAACAGTCGGGAGAGGAGTGGCATTA
ACACCATCCTTCATGAACTTAATCCACTGAGATCGGAAGAGCGGTTCAGCAG
GAATGCCGAGACCGCCATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 43)

8 (198bp) Insert: PhiX 79bp

IDT Exhibit 1018.0697

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGATAGTTTGACGGTTAATGCTGGTAATGGTGGTTTTCTTCATTGCAT
TCAGATGGATACATCTGTCAACGCCGCTAATCAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCGCCATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 44)

9 (179bp) Insert: PhiX 60bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTGAAACCAACATAAACATTATTGCCCGGCGTACCGAGGAAGGACGT
CAATAGTCACACAGTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC
GCCATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 45)

10 (198bp) Insert: PhiX 79bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACATCGT
CAACGTTATATTTTGATAGTTTGACGGTTAATAAGATCGGAAGAGCGGTTCA
GCAGGAATGCCGAGACCGCCATCGTATGCCGTCTTCTGCTTG (SEQ ID No. 46)

11 (151bp) Insert: PhiX 32bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGCCATCGTATGCCGTCTTCTGCTTG (SE
Q ID No. 47)

12 (173bp) Insert: PhiX 54bp
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTTTTGCTGATGAACTAAGTCAACCTCAGCACTAACCTTGCGAGTCAT
TTCTTTGAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCCATCG
TATGCCGTCTTCTGCTTG (SEQ ID No. 48)

13 (151bp) Insert: PhiX 32bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTAATAGACGCAACGCGAGCAGTAGACTCCTTCTAGATCGGAAGAGC
GGTTCAGCAGGAATGCCGAGACCGCCATCGTATGCCGTCTTCTGCTTG(SE
Q ID No. 49)

14 (168bp) Insert: PhiX 49bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGGAGGCCTCCAGCAATCTTGAACACTCATCCTTAATACCTTTCTTTT
TGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGCCATCGTATGC
CGTCTTCTGCTTG(SEQ ID No. 50)

15 (181bp) Insert: PhiX 62bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTACTCCTAAGCAGAAAACCTACCGCGCTTCGCTTGGTCAACCCCTCA
GCGGCAAAAATTAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAC
CGCCATCGTATGCCGTCTTCTGCTTG(SEQ ID No. 51)

16 (182bp) Insert: E.coli 63bp

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
CGATCTGAGCCGACATCGAGGTGCCAAACACCGCCGTCGATATGAACTCTT
GGGCGGTATCAGCCTGTTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA
GACCGCCATCGTATGCCGTCTTCTGCTTG(SEQ ID No. 52)

These results confirm that the library preparation method produces 2 libraries of 'sequenceable' DNA templates. Each library contained a plurality of genomic inserts, each of which was flanked by the two adaptors (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID No. 53) and AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGNNNTCGTATGCCGTCTTCTGCTT G(SEQ ID No. 54)), required for cluster formation and SBS

sequencing. The insert DNA from each of the libraries aligned to the correct reference, either BAC or PhiX, with a small amount of E.coli contamination in each library. The correct TAG ([redacted] for PhiX or [redacted] for BAC) was present in each library.

Example of Cluster preparation using the mixed PHiX and BAC libraries

Example 2): Acrylamide coating of glass chips

The solid supports used are typically 8-channel glass chips such as those provided by Silex Microsystems (Sweden). However, the experimental conditions and procedures are readily applicable to other solid supports.

Chips were washed as follows: neat Decon for 30 minutes, milliQ H_2O for 30 minutes, NaOH 1N for 15 minutes, milliQ H_2O for 30 minutes, HCl 0.1N for 15 minutes, milliQ H_2O for 30 minutes.

Polymer solution preparation:

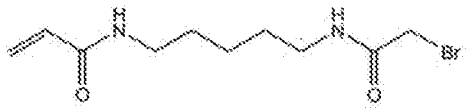
For 10 ml of 2% polymerisation mix.

- 10 ml of 2% solution of acrylamide in milliQ H2O
- 165 µl of a 100mg/ml *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA) solution in DMF (23.5 mg in 235µl DMF)
- 11.5 µl of TEMED
- 100 µl of a 50 mg/ml solution of potassium persulfate in milliQ H_2O (20mg in 400µl H_2O)

The 10 ml solution of acrylamide was first degassed with argon for 15 minutes. The solutions of BRAPA, TEMED and potassium persulfate were successively added to the

acrylamide solution. The mixture was then quickly vortexed and immediately used. Polymerization was then carried out for 1h 30 minutes at RT. Afterwards the channels were washed with milliQ H_2O for 30 minutes and filled with 0.1 M potassium phosphate buffer for storage until required.

Example 3) Synthesis of *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA) (1)



(1)

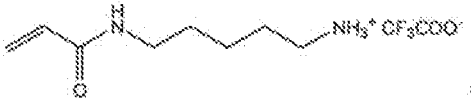
N-Boc-1,5-diaminopentane toluene sulfonic acid was obtained from Novabiochem. The bromoacetyl chloride and acryloyl chloride were obtained from Fluka. All other reagents were Aldrich products.



(2)

To a stirred suspension of *N*-Boc-1,5-diaminopentane toluene sulfonic acid (5.2 g, 13.88 mmol) and triethylamine (4.83 ml, 2.5 eq) in THF (120 ml) at 0°C was added acryloyl chloride (1.13 ml, 1 eq) through a pressure equalized dropping funnel over a one hour period. The reaction mixture was then stirred at room temperature and the progress of the reaction checked by TLC (petroleum ether:ethyl acetate 1:1). After two hours, the salts formed during the reaction were filtered off and the filtrate evaporated to dryness. The residue was purified by flash chromatography (neat petroleum ether followed by a gradient of ethyl acetate up to 60%) to

yield 2.56 g (9.98 mmol, 71 %) of product **2** as a beige solid. ^{1}H NMR (400 MHz, d_6-DMSO) : 1.20-1.22 (m, 2H, CH_2), 1.29-1.43 (m, 13H, tBu, 2xCH_2), 2.86 (q, 2H, J = 6.8 Hz and 12.9 Hz, CH_2), 3.07 (q, 2H, J = 6.8 Hz and 12.9 Hz, CH_2), 5.53 (dd, 1H, J = 2.3 Hz and 10.1 Hz, CH), 6.05 (dd, 1H, J = 2.3 Hz and 17.2 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 17.2 Hz, CH), 6.77 (t, 1H, J = 5.3 Hz, NH), 8.04 (bs, 1H, NH). Mass (electrospray+) calculated for $C_{13}H_{24}N_2O_3$ 256, found 279 (256+Na^+).



(**3**)

Product **2** (2.56g, 10 mmol) was dissolved in trifluoroacetic acid:dichloromethane (1:9, 100 ml) and stirred at room temperature. The progress of the reaction was monitored by TLC (dichloromethane:methanol 9:1). On completion, the reaction mixture was evaporated to dryness, the residue co-evaporated three times with toluene and then purified by flash chromatography (neat dichloromethane followed by a gradient of methanol up to 20%). Product **3** was obtained as a white powder (2.43 g, 9 mmol, 90%). ^{1}H NMR (400 MHz, D_2O): 1.29-1.40 (m, 2H, CH_2), 1.52 (quint., 2H, J = 7.1 Hz, CH_2), 1.61 (quint., 2H, J = 7.7 Hz, CH_2), 2.92 (t, 2H, J = 7.6 Hz, CH_2), 3.21 (t, 2H, J = 6.8 Hz, CH_2), 5.68 (dd, 1H, J = 1.5 Hz and 10.1 Hz, CH), 6.10 (dd, 1H, J = 1.5 Hz and 17.2 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 17.2 Hz, CH). Mass (electrospray+) calculated for $C_8H_{16}N_2O$ 156, found 179 (156+Na^+).

To a suspension of product **3** (6.12 g, 22.64 mmol) and triethylamine (6.94 ml, 2.2 eq) in THF (120 ml) was added

bromoacetyl chloride (2.07 ml, 1.1eq), through a pressure equalized dropping funnel, over a one hour period and at -60°C (cardice and isopropanol bath in a Dewar). The reaction mixture was then stirred at room temperature overnight and the completion of the reaction was checked by TLC (dichloromethane:methanol 9:1) the following day. The salts formed during the reaction were filtered off and the reaction mixture evaporated to dryness. The residue was purified by chromatography (neat dichloromethane followed by a gradient of methanol up to 5%). 3.2 g (11.55 mmol, 51 %) of the product **1** (BRAPA) were obtained as a white powder. A further recrystallization performed in petroleum ether:ethyl acetate gave 3g of the product **1**. ^{1}H NMR (400 MHz, d_6-DMSO) : 1.21-1.30 (m, 2H, CH_2), 1.34-1.48 (m, 4H, 2xCH_2), 3.02-3.12 (m, 4H, 2xCH_2), 3.81 (s, 2H, CH_2), 5.56 (d, 1H, J = 9.85 Hz, CH), 6.07 (d, 1H, J = 16.9 Hz, CH), 6.20 (dd, 1H, J = 10.1 Hz and 16.9 Hz, CH), 8.07 (bs, 1H, NH), 8.27 (bs, 1H, NH). Mass (electrospray+) calculated for $C_{10}H_{17}BrN_2O_2$ 276 or 278, found 279 (278+H^+), 299 (276+Na^+).

Example 4) Grafting primers onto surface of SFA coated chip

An SFA coated chip is placed onto a modified MJ-Research thermocycler and attached to a peristaltic pump. Grafting mix consisting of 0.5 µM of a forward primer and 0.5 µM of a reverse primer in 10 mM phosphate buffer (pH 7.0) is pumped into the channels of the chip at a flow rate of 60 µl/minutes for 75 s at 20 °C. The thermocycler is then heated to 51.6 °C, and the chip is incubated at this temperature for 1 hour. During this time, the grafting mix undergoes 18 cycles of pumping: grafting mix is pumped in at 15 µl/minutes for 20 s, then the solution is pumped back and

forth (5 s forward at 15 µl/minutes, then 5 s backward at 15 µl/minutes) for 180 s. After 18 cycles of pumping, the chip is washed by pumping in 5xSSC/5mM EDTA at 15 µl/minutes for 300 s at 51.6 °C. The thermocycler is then cooled to 20 °C.

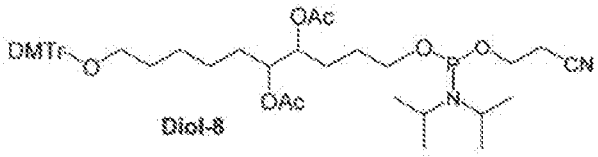
The primers are typically 5'-phosphorothioate oligonucleotides incorporating any specific sequences or modifications required for cleavage. Their sequences and suppliers vary according to the experiment for which they are to be used, and in this case were complementary to the 5'-ends of the template duplex. The DNA sequence used in this process was the pool of the two libraries, which have ends complementary to the grafted primers. The library mix was denatured using sodium hydroxide treatment followed by snap dilution as described.

For some of the experiments detailed, the amplified clusters contained a diol linkage in one of the grafted primers. Diol linkages can be introduced by including a suitable linkage into one of the primers used for solid-phase amplification. Synthesis of the diol phosphoramidite is described in Example 4 below. Products containing such diol linkages can be cleaved using periodate and propanolamine as described, and the resulting single stranded polynucleotides hybridised as described.

The grafted primers contain a sequence of T bases at the 5'-end to act as a spacer group to aid linearisation and hybridization. The sequences of the two primers grafted to the chip are as follows:

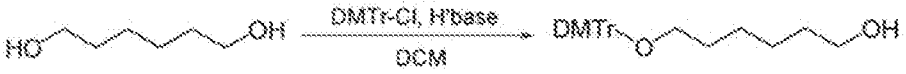
P5diol: 5' PS-TTTTTTTTTT-diol-AATGATACGGCGACCACCGA (SEQ ID No. 55)
P7: 5' PS-TTTTTTTTTTCAAGCAGAAGACGGCATACGA (SEQ ID No. 56)

Example 5: Preparation of diol-phosphoramidite for DNA coupling



Diol-8

Step 1:



Diol-1

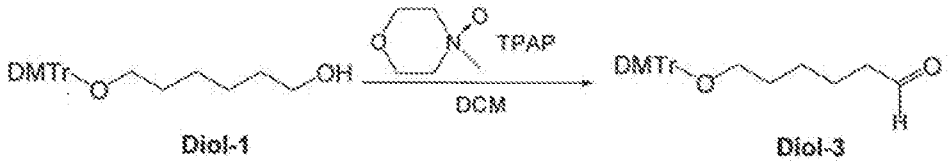
1,6-Hexanediol (Sigma Aldrich 99 %) (14.6 g, 124 mmol), *N,N*-diisopropylethylamine (Hünig's base; Sigma Aldrich; redistilled) (21.6 mL, 124 mmol) is dissolved in anhydrous DCM/ DMF (250/ 50 mL) under N_2. The solution is cooled to 0°C and the first portion of 4, 4'-dimethoxytrityl chloride (DMTr-Cl; Sigma-Aldrich 95 %) (10.5 g, 31 mmol) is added. The reaction mixture is then warmed to room temperature. After stirring for 1 h, the reaction mixture is cooled to 0°C again and the second portion of DMTr-Cl (10.5 g, 31 mmol) is added and then allowed to stir at room temperature for another 2 hours. TLC (EtOAc: petroleum ether 4:6) analysis indicates ca. 95% consumption of starting material derivative (DMTr-OH). The reaction is concentrated under reduced pressure and Aq. $NaHCO_3$ (sat.) solution (500 mL) is poured into the residue. The resulting mixture is extracted with petroleum ether/ EtOAc (2:1) (3 x 1000 mL). The

combined organic layers are dried over $MgSO_4$, and concentrated under vacuum. The residue is co-evaporated with xylene (2 x 100 mL) to remove DMF. The reaction mixture, is pre-absorbed on silica gel and subjected to flash chromatography using solvents containing 1 % Et_3N petroleum ether to petroleum ether/ EtOAc (7:3) as eluent. The yield of pale yellow oil is 16.58 g, 64 %, with a further 7.8g (17%) of the bis-tritylated by-product.

TLC: R_f: 0.35 (diol-1); R_f: 0.7 (bis-tritylated by-product) (petroleum ether/ EtOAc 6:4).

^1H NMR (400 MHz, $CDCl_3$): δ 1.32-1.44 (m, 4H, 2x CH_2), 1.54-1.68 (m, 4H, 2x CH_2), 3.06 (t, J = 6.6 Hz, 2H, CH_2O), 3.62-3.68 (m, 2H, CH_2OH), 3.81 (s, 6H, 2x MeO), 6.83-6.85 (m, 4H, Ph), 7.24-7.35 (m, 7H, Ph), 7.45-7.47 (m, 2H, Ph).

Step 2:



Diol-1 Diol-3

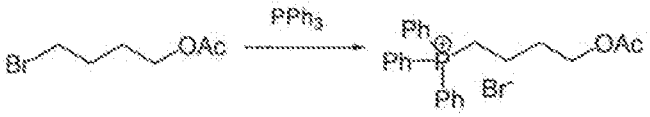
To a solution of Diol-1 (16.6 g, 39.5 mmol) in anhydrous DCM (200 mL), tetrapropylammonium perruthenate (TPAP; Sigma Aldrich 97 %) (277 mg, 0.79 mmol) is added under N_2 atmosphere. The solution is cooled to 0 °C and *N*-methylmopholine *N*-oxide (Sigma Aldrich 97 %) (2.7 g, 23 mmol) is added. The reaction is warmed to room temperature. After 1 hour, the other three portions of *N*-methylmopholine *N*-oxide (3 x 2.0 g, 51.2 mmol) are added within a period of

four hours. TLC (EtOAc: petroleum ether 4:6) indicates the reaction goes to completion. The reaction is quenched with aq. $NaHCO_3$ (sat.) (1000 mL) and extracted to CH_2Cl_2 (4 x 1000 mL). The combined organic layers are dried over $MgSO_4$. The solution is concentrated under reduced pressure. Diol-3, 9.9 g, 60 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (6:4) as eluent, as a pale yellow oil.

TLC: R_f: 0.7 (petroleum ether/ EtOAc 6:4).

^1H NMR (400MHz, $CDCl_3$): δ 1.30-1.37 (m, 2H, CH_2), 1.48-1.57 (m, 4H, 2x CH_2), 2.34 (td, J = 1.7 and 7.4 Hz, 2H, CH_2CHO), 2.97 (s, 2H, CH_2O), 3.72 (s, 6H, 2x MeO), 6.73-6.76 (m, 4H, Ph), 7.10-7.26 (m, 7H, Ph), 7.34-7.36 (m, 2H, Ph), 9.67 (t, J = 1.7, 1H, CHO).

Step 3:



Diol-2

A solution of triphenylphosphine (Sigma-Aldrich 99 %, ReagentPlus™). (39.3 g, 150 mmol) and 4-bromobutyl acetate (Sigma-Aldrich)(26 mL, 180 mmol) in anhydrous toluene (300 mL) is heated under reflux for 36 hours under N_2 in an oil-bath (140 °C). During the reflux, oil is precipitated out. The reaction mixture is cooled to room temperature. TLC (petroleum ether/ EtOAc 7:3) analysis of the toluene

solution showed that there is still triphenylphosphine (R_f: 0.8) left. The supernatant is decanted into another round-bottomed flask and concentrated down to the approximate volume of 30 mL. The solution is heated under reflux again for another 12 hours. The supernatant is decanted. The portions of oil are combined together, dissolved in water (500 mL) and extracted with EtOAc (2 x 500 mL). The combined organic layers are back-extracted with water (150 mL). Two lots of aqueous layers are combined, and evaporated under reduced pressure. The resulting residue is co-evaporated with acetonitrile (2 x 100 mL) to give 78.4 g, 95 % yield of a pale yellow oil. NMR indicates that the product is pure, and is used for the next step reaction without further purification.

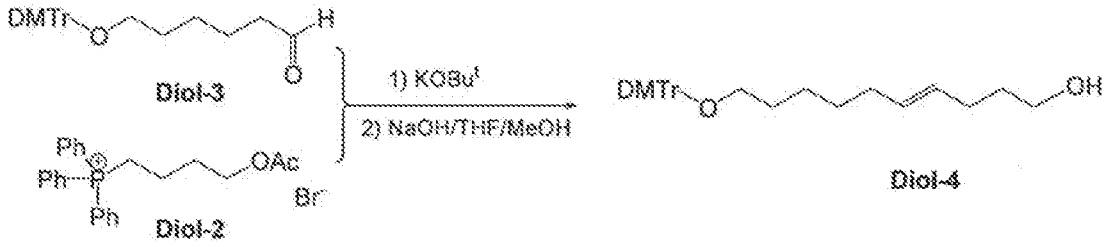
TLC: R_f: 0.0 (petroleum ether/ EtOAc 7:3).

^{1}H NMR (400 MHz, $CDCl_3$): δ 1.63-1.73 (m, 2H, CH_2), 1.94 (s, 3H, 2x CH_3), 2.06-2.16 (m, 2H, CH_2), 3.97-4.05 (m, 2H, CH_2P), 4.11 (t, J = 6.0, 2H, CH_2O), 7.69-7.95 (m, 15H, Ph).

^{31}P NMR (162 MHz, $CDCl_3$): 25.9 ppm.

Mass spec details: LC-MS (Electrospray positive): (M^+) 377.

Step 4:



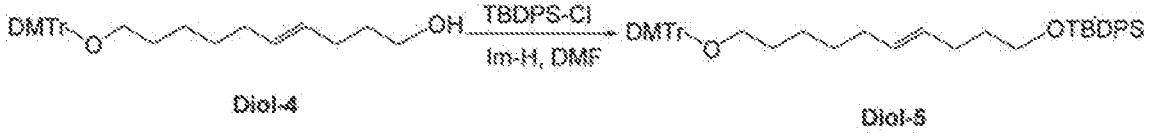
Diol-2 (10.34 g, 22.7 mmol) is weighed into a round-bottomed flask and dissolved with DCM (20 mL). The solution is then evaporated under reduced pressure until it gives a white foam. The flask is then subjected to high vacuum for 24 h. To this flask, anhydrous THF (180 mL) is added under N_2. The resulting suspension is cooled to -78 °C with an acetone-dry ice bath. With vigorous stirring, $KOBu^t$ (3.3 g, 29.5 mmol) is added under N_2. Slowly the colour of the suspension turns orange, and white solids are gradually precipitated out. To this suspension, a solution of diol-3 (dried at 60 °C under high vacuum for 1 h before the reaction),(9.5 g, 22.7 mmol) in THF (50 mL) is added drop wise over half an hour. Acetone-dry ice bath is then removed. The reaction mixture is slowly warmed to room temperature and stirred for another hour. The colour of the reaction mixture turns yellow after the addition of diol-3. The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. NaCl (sat.) (800 mL). The aqueous layer is extracted with an additional DCM (2 x 800 mL). The organic extractions are combined, dried over $MgSO_4$, filtered, and evaporated under reduced pressure to give yellow oil. The oil is dissolved in THF/MeOH (125/100 mL) and cooled to 0°C. To this solution, NaOH (1M in H_2O, 25 mL) is added. After allowing the reaction to stir for 1 hour, TLC analysis

indicates full consumption of starting material. The reaction mixture is neutralized with acetic acid (1.5 mL). The reaction mixture is concentrated down under reduced pressure. The resulting residue is partitioned between DCM (800 mL) and aq. $NaHCO_3$ (sat.) (800 mL). The aqueous layer is extracted with additional DCM (2 x 800 mL). The organic extractions are combined, dried over $MgSO_4$, filtered, and evaporated to give a pale yellow oil. Diol-4, 6.45 g, 60 % is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (5:5) as eluent, as a light yellow oil.

TLC: R_f = 0.45 (petroleum ether/ EtOAc 6:4).

^1H NMR (400 MHz, $CDCl_3$) δ 1.24-1.32 (m, 4H, 2x CH_2), 1.54-1.57 (m, 4H, 2x CH_2), 1.93-1.96 (m, 2H, CH_2), 2.02-2.07 (m, 2H, CH_2), 2.96 (t, J = 6.6 Hz, 2H, CH_2O), 3.54-3.59 (m, 2H, CH_2OH), 3.72 (s, 6H, 2x MeO), 5.29-5.32 (m, 2H, 2x =CH), 6.73-6.77 (m, 4H, Ph), 7.11-7.27 (m, 7H, Ph), 7.36-7.38 (m, 2H, Ph).

Step 5:



Diol-4 Diol-5

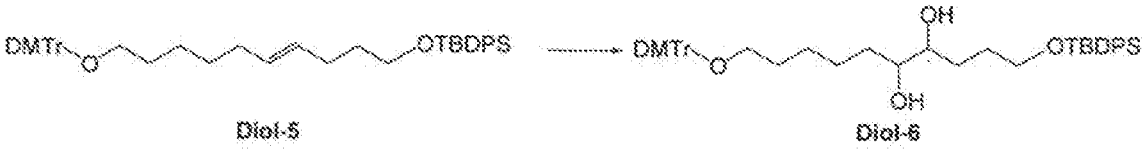
To a solution of Diol-4 (5.68 g, 12 mmol) and imidazole (Sigma Aldrich, 99 %), (1.63 g, 24 mmol) in anhydrous DMF (100 mL), t-butyldiphenylsilyl chloride (Sigma Aldrich, 98

%), (4.05 mL, 15.6 mmol) is added drop wise under N_2 atmosphere at room temperature. The reaction is stirred for 1 hour. TLC (petroleum ether/ EtOAc 8:2) indicates that the starting material is fully consumed. A saturated aq. $NaHCO_3$ solution (500 mL) is added to quench the reaction. The resulting mixture is extracted with petroleum ether/EtOAc (2:1) (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give a yellow oil. Diol-5, 8.14 g, 95 % is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (9:1) as eluent, as a colourless oil.

TLC: R_f = 0.7 (petroleum ether : EtOAc 8:2).

^1H NMR (400 MHz, $CDCl_3$): δ 0.97 (s, 9H, 3xMe), 1.19-1.30 (m, 4H, 2x CH_2), 1.48-1.55 (m, 4H, 2x CH_2), 1.91-1.95 (m, 2H, CH_2), 2.01-2.06 (m, 2H, CH_2), 2.95 (t, J = 6.6 Hz, 2H, CH_2O), 3.58 (t, J = 6.3 Hz, 2H, CH_2O), 3.70 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.72-6.75 (m, 4H, Ph), 7.11-7.37 (m, 15H, Ph), 7.57-7.60 (m, 4H, Ph).

Step 6:



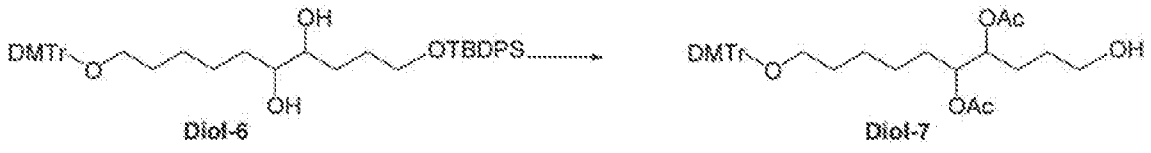
A mixture of diol-5 (9.27 g, 13 mmol), AD-mix-α (Sigma Aldrich), (18.2 g), methanesulfonamide (Sigma Aldrich, 97 %), (1.23 g, 13 mmol), t-BuOH (65 mL) and water (65 mL) is stirred together vigorously at 55 °C for 14 h. The TLC

analysis (petroleum ether : EtOAc 6:4) indicates ca. 95 % consumption of the starting material. The reaction mixture is cooled to room temperature, treated with sodium sulfite (15.3 g, 12 mmol), then further stirred for 30 minutes. A saturated aq. $NaHCO_3$ solution (500 mL) is added to the reaction. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give yellow oil. Diol-6, 7.96 g, 82 %, is isolated by flash chromatography (silica gel, Fluka, 70-230 mesh) using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (1:1) as elutant, as a white solid.

TLC: R_f = 0.3 (petroleum ether : EtOAc 6:4).

^1H NMR (400 MHz, $CDCl_3$) δ 1.07 (s, 9 H, 3xMe), 1.41-1.7 (m, 12 H, 6x CH_2), 1.94 (d, J = 4.3 Hz, 1H, OH), 2.94-2.95 (m, 1H, OH), 3.06 (t, J = 6.6 Hz, 2H, CH_2O), 3.61-3.63 (m, 2 H, 2x C*H*OH), 3.73 (t, J = 5.6 Hz, 2H, CH_2O), 3.81 (s, 6H, 2x MeO), 5.24-5.27 (m, 2H, 2x =CH), 6.82-6.85 (m, 4H, Ph), 7.21-7.47 (m, 15 H, Ph), 7.57-7.60 (m, 4 H, Ph).

Step 7:



To a solution of diol-6 (7.96 g, 13 mmol) and DMAP (Sigma-Aldrich ReagentPlus℠, 99%). (260 mg, 2.13 mmol) in a

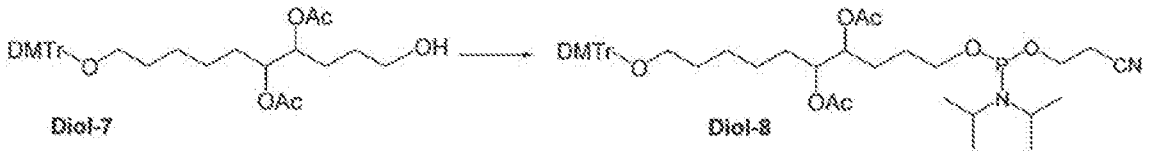
mixture of pyridine (15 mL) and DCM (30 mL), acetic anhydride (Fluka 99%), (2.5 mL, 26.68 mmol) is added at room temperature. TLC analysis (petroleum ether:EtOAc 6:4) indicates full consumption of the starting material after 1 h. The reaction is quenched by saturated aq. $NaHCO_3$ solution (500 mL). After 5 minutes, the mixture is extracted with DCM (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated. The residue is co-evaporated with toluene (2 x 100 mL). The resulting yellow oil is subjected to a plug of silica gel (50 g, Fluka, 70-230 mesh) to remove DMAP using eluents containing 0.1% Et_3N from petroleum ether to petroleum ether/ EtOAc (7:3) (250 mL each). The combined fractions of product are concentrated to dryness. The resulting colourless oil is dissolved in THF (100 mL) and treated with TBAF (Sigma-Aldrich; 5% wt water), (1 M in THF, 15 mL) at 0 °C. The reaction solution is slowly warmed to room temperature and stirred for 2 hours. TLC analysis (petroleum ether:EtOAc 6:4) indicates that desilylation is completed. The volatile solvent (THF) is evaporated under reduced pressure at low temperature. A saturated aq. $NaHCO_3$ solution (500 mL) is added to the residue. The resulting mixture is extracted with EtOAc (3 x 500 mL). The organic layers are combined, dried over $MgSO_4$, filtered, and evaporated to give yellow oil. Diol-7, 4.2 g, 66 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (1:1) as eluent, as a white solid.

TLC: R_f = 0.45 (petroleum ether : EtOAc 1:1).

^1H NMR (400 MHz, $CDCl_3$) δ 1.29-1.33 (m, 4H, 2x CH_2), 1.47-1.63 (m, 8 H, 4x CH_2), 1.99, 2.01 (2s, 6H, 2 MeC(O)), 3.00

(t, J = 6.5 Hz, 2H, CH_2O), 3.60-3.64 (m, 2 H, CH_2O), 3.75 (s, 6H, 2x MeO), 4.92-4.97 (m, 2H, 2x C*H*OAc), 6.76-6.80 (m, 4H, Ph), 7.15-7.29 (m, 7 H, Ph), 7.38-7.40 (m, 2 H, Ph).

Step 8:



To a solution of diol-7 (2.08 g, 3.5 mmol) and diisopropylethylamine (Sigma Aldrich), (1.53 ml, 8.75 mmol) in DCM (17 mL), 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (1.0 g, 4.2 mmol) is added drop wise at room temperature under N_2. After stirring for 1 hour, TLC analysis (petroleum ether:EtOAc 4:6) indicates the full consumption of the starting material. The solvent (THF) is concentrated under reduced pressure. The resulting residue is subjected to chromatography directly. Diol-8, 2.5 g, 90 %, is isolated by flash chromatography using solvents containing 1% Et_3N from petroleum ether to petroleum ether/ EtOAc (1:1) as eluent, as a colourless syrup.

TLC: R_f = 0.55 (petroleum ether : EtOAc 4:6).

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.09, 1.10, 1.11, 1.12 (4 s, 12 H, $N(CHMe_2)_2$), 1.26-1.31 (m, 4H, 2x CH_2), 1.45-1.56 (m, 8 H, 4x CH_2), 1.95, 1.969, 1.971, 1.98 (4 s, 6H, 2 MeCO), 2.56 (t, J = 6.5 Hz, 2H, CH_2CN), 2.95 (t, J = 6.5 Hz, 2H, CH_2O), 3.49-3.55 (m, 4 H, CH_2O), 3.72 (s, 6H, 2x MeO), 4.89-4.92 (m, 2H,

2x CHOAc), 6.74-6.76 (m, 4H, Ph), 7.13-7.25 (m, 7 H, Ph), 7.34-7.37 (m, 2 H, Ph).

^{31}P NMR (162 MHz, $CDCl_3$): 148.67, 148.69 ppm.

Example 6) Preparation of clusters by isothermal amplification:

Step 1: Hybridisation and amplification

The DNA sequence used in the amplification process is the mixture of the two libraries prepared in Example 1, which have ends complementary to the grafted primers. The duplex DNA (1 nM) is denatured using 0.1 M sodium hydroxide treatment followed by snap dilution to the desired 0.2-2 pM 'working concentration' in 'hybridization buffer' (5 x SSC / 0.1% Tween).

Surface amplification was carried out by isothermal amplification using an MJ Research thermocycler, coupled with an 8-way peristaltic pump Ismatec IPC ISM931 equipped with Ismatec tubing (orange/yellow, 0.51 mm ID). A schematic of the instrument is shown in figure 10.

The single stranded template (denatured as indicated above) is hybridised to the grafted primers immediately prior to the amplification reaction, which thus begins with an initial primer extension step rather than template denaturation. The hybridization procedure begins with a heating step in a stringent buffer to ensure complete denaturation prior to hybridisation. After the hybridization, which occurs during a 20 minutes slow cooling step, the flowcell was washed for 5 minutes with a wash buffer (0.3 x SSC / 0.1% Tween).

A typical amplification process is presented in Table 1, which details the flow volumes per channel.

Table 1. Template Hybridization and 1st Extension

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump Hybridization pre-mix	20	120	60	120
2	Pump Hybridization mix	98.5	300	15	75
3	Remove bubbles	98.5	10	100	16.7
4	Stop flow and hold T	98.5	30	static	0
5	Slow cooling	98.5-40.2	19.5 min	static	0
6	Pump wash buffer	40.2	300	15	75
7	Pump amplification pre-mix	40.2	200	15	50
8	Pump amplification mix	40.2	75	60	75
9	First Extension	74	90	static	0
10	cool to room temperature	20	0	static	0

Table 2. Isothermal Amplification

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
(1) This sequence 35 times	Pump Formamide	60	75	60	75
	Pump Amplification pre-mix	60	75	60	75
	Pump Bst mix	60	95	60	95
	Stop flow and hold T	60	180	static	0
2	Pump wash buffer	60	120	60	120

Details are as follows: Hybridisation pre mix (buffer) = 5 x SSC / 0.1% Tween; Hybridisation mix = 0.1 M hydroxide DNA sample, diluted in hybridisation pre mix; Wash buffer = 0.3 x SSC / 0.1% Tween; Amplification pre mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8; Amplification mix = 2 M

betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus 200 µM dNTP's and 25 units/mL of Taq polymerase (NEB Product ref M0273L); Bst mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus 200 µM dNTP's and 80 units/mL of Bst polymerase (NEB Product ref M0275L)

Step 2: Blocking extendable 3'-OH groups

To prepare the blocking pre-mix, 1530 µL of water and 170 µL of 10X blocking buffer (NEB buffer 4; product number B7004S) are mixed for a final volume of 1700 µL. To prepare the blocking mix, 1065.13 µL of blocking pre-mix, 21.12 µL of 125 µM ddNTP mix, and 13.75 µL of TdT terminal transferase (NEB; part no M0252S) are mixed for a final volume of 1100 µL.

To block the nucleic acid within the clusters formed in the flow cell channels, the computer component of the instrumentation flows the appropriate blocking buffer through the flow cell, and controls the temperature as shown in the exemplary embodiments below in Table 3.

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump Blocking pre-mix	20	200	15	50
2	Pump Blocking mix	37.7	300	15	75
3	Stop flow and hold T	37.7	20	static	0
4	Cyclic pump Blocking mix and	37.7	8 x (20+180)	15 / static	45

	wait				
5	Pump wash buffer	20	300	15	75

Example 7: Linearisation and hybridization of a sequencing primer

To prepare the linearization mix, 1429 µL of water, 64 mg of sodium periodate, 1500 µL of formamide, 60 µL of 1M Tris pH8, and 6011.4 µL of 3-aminopropanol are mixed for a final volume of 3 mL. The periodate is first mixed with the water while the Tris is mixed with the formamide. The two solutions are then mixed together and the 3-aminopropanol is added to that mixture.

To linearize the nucleic acid within the clusters formed within the flow cell channels, 300 µL per channel of linearisation mix is flowed in at 15 µL/minutes at 20 °C; followed by 75 µL of water at the same flow rate.

To prepare the primer mix, 895.5 µL of hybridization buffer and 4.5 µl of sequencing primer (100 µM) are mixed to a final volume of 900 µL. The sequence of the sequencing primer used in the first reaction was:

5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC (SEQ ID No. 57)

To denature the nucleic acid within the clusters and to hybridize the sequencing primer, the following reagents are flowed through the cell as shown in Table 4:

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump 0.1 M NaOH	20	300	15	75
2	Pump TE	20	300	15	75
3	Pump Primer mix	20	300	15	75

4	Hold at 60C	60	900	0	0
5	Pump wash buffer	40.2	300	15	75

Example 8: Sequencing from the target fragment

Sequencing of the clusters from the above illustrative protocol was carried out using modified nucleotides prepared as described in International patent application WO 2004/018493, and labeled with four spectrally distinct fluorophores, as described in US application number 60/801270; filed May 18th 2006. Sequencing of clusters is described in more detail in patent WO06064199. The contents of the above-listed three documents are incorporated herein by reference in their entireties.

A mutant 9°N polymerase enzyme (an exo- variant including the triple mutation L408Y/Y409A/P410V and C223S) was used for the nucleotide incorporation steps.

Incorporation mix, Incorporation buffer (50 mM Tris-HCl pH 8.0, 6 mM MgSO4, 1 mM EDTA, 0.05% (v/v) Tween -20, 50mM NaCl) plus 110nM YAV exo- C223S, and 1μM each of the four labeled modified nucleotides, was applied to the clustered templates, and heated to 45°C.

Templates were maintained at 45°C for 30 minutes, cooled to 20°C and washed with Incorporation buffer, then with 5x SSC/0.05% Tween 20. Templates were then exposed to Imaging buffer (100mM Tris pH7.0, 30mM NaCl, 0.05% Tween 20, 50mM sodium ascorbate, freshly dissolved). Templates were scanned in 4 colors at room temperature. Templates were then exposed to sequencing cycles of Cleavage and Incorporation as follows:

Cleavage: Prime with Cleavage buffer (0.1M Tris pH 7.4, 0.1M NaCl and 0.05% Tween 20), 125 µL/channel; 60 µL/minutes.

Heat to 60 °C.

Treat the clusters with Cleavage mix (100 mM TCEP in Cleavage buffer), 75 µL/channel; 60 µL/minutes.

Wait for a total of 15 minutes in addition to pumping fresh cleavage mix, 25 µL/channel; 60 µL/minutes, every 4 minutes.

Cool to 20 °C.

Wash with Enzymology buffer.

Wash with 5XSSC/0.05% Tween 20.

Prime with Imaging buffer.

Scan in 4 colors at RT.

Incorporation: Prime with Incorporation buffer, 125 µL/channel; 60 µL/minutes, Heat to 60°C

Treat with Incorporation mix, 75 µL/channel; 60 µL/minutes.

Wait for a total of 15 minutes in addition to pumping fresh Incorporation mix, 25 µL/channel; 60 µL/minutes, every 4 minutes.

Cool to 20 °C.

Wash with Incorporation buffer, 75 µL/channel; 60 µL/minutes.

Wash with 5XSSC/0.05% Tween 20, 75 µL/channel; 60 µL/minutes

Prime with Imaging buffer, 100 µL/channel; 60 µL/minutes

Scan in 4 colors at RT.

Repeat the process of Incorporation and Cleavage for as many cycles as required.

Incorporated nucleotides were detected using a Total Internal Reflection based fluorescent CCD imaging apparatus described in "Systems and Devices for Sequence by Synthesis Analysis," USSN 60/788,248, filed March 31, 2006.

Example 9) Sequencing the tag

The first sequencing primer can be removed, and a second one applied using the same protocol as described in Example 7. The sequence of the second sequencing primer was as follows:

5'-AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG (SEQ ID No. 58)

The sequencing cycles described in Example 8 can be repeated to sequence the tag. In this case, only three cycles of sequencing were performed to analyse the tag. The data in Figure 4 show that all the clusters sequenced aligned against a known sample (PhiX, BAC or E. coli contamination), and that for the samples that aligned correctly, 98% showed the correct tag. The 0.1% errors derive from mis-alignments of the target read, rather than errors in the tag sequences.

Example 10) 12-plex paired reads from a library of fragments

Step 1) Grafting primers onto surface of SFA coated chip

IDT Exhibit 1018.0721

An SFA coated chip is placed onto a modified MJ-Research thermocycler and attached to a peristaltic pump. Grafting mix consisting of 0.5 µM of a forward primer and 0.5 µM of a reverse primer in 10 mM phosphate buffer (pH 7.0) is pumped into the channels of the chip at a flow rate of 60 µl/minutes for 75 s at 20 °C. The thermocycler is then heated to 51.6 °C, and the chip is incubated at this temperature for 1 hour. During this time, the grafting mix undergoes 18 cycles of pumping: grafting mix is pumped in at 15 µl/minutes for 20 s, then the solution is pumped back and forth (5 s forward at 15 µl/minutes, then 5 s backward at 15 µl/minutes) for 180 s. After 18 cycles of pumping, the chip is washed by pumping in 5xSSC/5mM EDTA at 15 µl/minutes for 300 s at 51.6 °C. The thermocycler is then cooled to 20 °C.

The primers are typically 5'-phosphorothioate oligonucleotides incorporating any specific sequences or modifications required for cleavage. Their sequences and suppliers vary according to the experiment for which they are to be used, and in this case were complementary to the 5'-ends of the template duplex. The DNA sequence used in this process was the pool of the twelve libraries, which have ends complementary to the grafted primers. The library mix was denatured using sodium hydroxide treatment followed by snap dilution as described.

The grafted primers contain a sequence of T bases at the 5'-end to act as a spacer group to aid linearisation and hybridization. The sequences of the three primers grafted to the chip are as follows:

Oligo A: 5'-PS-TTTTTTTTTTAATGATACGGCGACCACCGAGAUCTACAC-3'
where U = 2-deoxyuridine; (SEQ ID No. 59)

Oligo B: 5'-PS-TTTTTTTTTTCAAGCAGAAGACGGCATAC**Goxo**A-3', where **Goxo** = 8-oxoguanine). (SEQ ID No. 60)

Step 2) Preparation of clusters by isothermal amplification:

The DNA sequence used in the amplification process is the mixture of the twelve libraries prepared in Example 1, which have ends complementary to the grafted primers. The duplex DNA (1 nM) is denatured using 0.1 M sodium hydroxide treatment followed by snap dilution to the desired 0.2-2 pM 'working concentration' in 'hybridization buffer' (5 x SSC / 0.1% Tween).

Surface amplification was carried out by isothermal amplification using a commercially available Solexa/Illumina cluster station as described in WO08002502. The cluster station is essentially a hotplate and a fluidics system for controlled delivery of reagents to a flow cell.

The single stranded template (denatured as indicated above) is hybridised to the grafted primers immediately prior to the amplification reaction, which thus begins with an initial primer extension step rather than template denaturation. The hybridization procedure begins with a heating step in a stringent buffer to ensure complete denaturation prior to hybridisation. After the hybridization, which occurs during a 20 minute slow cooling step, the flowcell was washed for 5 minutes with a wash buffer (0.3 x SSC / 0.1% Tween).

During template hybridization and first extension, a number of solutions/buffers are typically employed, e.g., a solution comprising the DNA samples, a hybridization buffer (5 x SSC / 0.1% tween), a wash buffer (0.3 x SSC / 0.1%

tween), a 2M sodium hydroxide solution, a cluster buffer (200 mM Tris, 100 mM Ammonium Sulfate, 20 mM Magnesium sulfate, 1% Triton, 1.3% DMSO, pH 8.8); an amplification additive (5 M betaine), DNA polymerase, and 10 mM dNTP mix.

To prepare the hybridization mixes, a 0.2 ml strip sample tube and the hybridization buffer are pre-chilled. Using 1.7 ml Eppendorf tube(s), the DNA template(s) are then diluted to 1nM in buffer EB (Qiagen). 1 µL of 2 M NaOH is added to 19 µL of template, vortexed briefly and incubated for 5 minutes at room temperature to denature the DNA template into single strands. The denatured DNA is diluted to working concentration (0.2-2 pM) in pre-chilled hybridization buffer (e.g. for 1 mL of 1pM Hybridization mix, 1 µL of denatured DNA is diluted into 1 mL of pre-chilled hybridization buffer). The volume required depends on the number of channels used - at least 120 µL of hybridization mix per channel is optionally used. Thus, 1 mL of hybridization mix is enough for 8 channels. The samples are vortexed briefly, spun down and aliquoted into the pre-chilled 0.2 ml strip tubes (with no bubbles in the bottom of the tubes) and used immediately.

To prepare the Amplification pre-mix (of sufficient volume for the first extension and 35 cycles of isothermal amplification), 35 mL of H_2O (milliQ), 7 mL of Cluster buffer (200 mM Tris, 100 mM Ammonium Sulfate, 20 mM Magnesium sulfate, 1% Triton, 1.3% DMSO, pH 8.8), and 28 mL of Amplification additive (5 M betaine solution) are mixed to achieve a final volume of 70 mL.

To prepare the first extension Taq mix, 780 µL of Amplification pre-mix, 16 µL of 10 mM dNTPs, and 4 µL of *Taq*

DNA polymerase are mixed together for a final volume of 800 µl.

A typical amplification process is detailed in the following table (Table 1), detailing the flow volumes per channel, controlled automatically by the computer component of the invention.

Table 1. Template hybridization and first extension.

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
1	Pump Hybridization pre-mix	20	120	60	120
2	Pump Hybridization mix	96	300	15	75
3	Remove bubbles	96	6	100	10
4	Stop flow and hold T	96	30	static	0
5	Slow cooling	96-40	1120	static	0
6	Pump wash buffer	40	300	15	75
7	Pump amplification pre-mix	40	280	15	70
8	Pump amplification mix	40	96	60	96
9	First Extension	74	90	static	0
10	cool to room temperature	20	0	static	0

Isothermal amplification at 60 °C using formamide as denaturant

The copied DNA can be isothermally amplified into clusters at 60 °C using formamide as a denaturant. The isothermal amplification (including both temperature control

and reagent control) is overseen by the computer component. Table 2 gives outlines of exemplary script controls. After the isothermal amplification, and optional washing step occur, the nucleic acid of the clusters is ready to be linearized (*see below*).

Table 2. Isothermal amplification

Step	Description	T (°C)	Time (sec)	Flow rate (µl/min)	Pumped V (µl)
(1) This sequence 35 times	Pump Formamide	60	56	30	28
	Pump Amplification pre-mix	60	56	30	28
	Pump Bst mix	60	72	30	36
2	Pump wash buffer	60	280	30	140
3	Pump Storage Buffer	20	380	15	95

Wash buffer = 0.3 x SSC / 0.1% Tween

Amplification pre mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8

Bst mix = 2 M betaine, 20 mM Tris, 10 mM Ammonium Sulfate, 2 mM Magnesium sulfate, 0.1% Triton, 1.3% DMSO, pH 8.8 plus 200 µM dNTPs and 80 units/mL of Bst polymerase (NEB Product ref M0275L)

Storage Buffer = 5X SSC.

Step 3) Preparation of clusters for first sequencing read

The preparation for read one was performed on the Illumina cluster station. All volumes used in the protocol were 95 ul per lane unless otherwise stated. Linearisation of A-type

surface immobilised oligonucleotides was achieved by incubation with USER enzyme mix (cocktail of Uracil DNA Glycosylase and Endonuclease VIII, NEB #M5505, 10 U/ml, 10 mM KCl, 20 mM Tris pH 8.8, 10 mM (HN4)2SO4, 2 mM MgSO4, 0.1% Triton X-100, 37 °C, 30 minutes). All exposed 3'-OH termini of DNA, either from the extended template or unextended surface oligonucleotides were blocked by dideoxy chain termination using a cocktail of terminal transferase (0.25 U/µl) and a modified polymerase (SBS polymerase as described below) (0.015 mg/ml, 100 µM ddNTP, 50 mM tris, 50 mM NaCl, 6 mM MgSO4, 1 mM EDTA, 0.05 % Tween 20). Blocking was achieved in a two stage protocol, initial incubation at 37 °C for 30 minutes followed by a ramping to 60 °C and incubating the flowcell for a further 15 minutes). Linearised and blocked clusters were washed with 0.3X SSC and storage buffer prior to denaturation with 0.1N NaOH. Denatured clusters were neutralised with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and washed with storage buffer. The read 1 specific sequencing primer

(5'- ACACTCTTTCCCTACACGACGCTCTTCCGATC -3' (SEQ ID No. 61), 0.5 µM in hybridisation buffer) was annealed to the clusters by incubation at 60 °C for 15 minutes, followed by a 0.3X SSC wash at 40 °C (ramp rate 1 °C / sec). The flowcell was finally flushed with storage buffer (at 20 °C). Processed flowcells were transferred to the Illumina Genome Analyser for sequencing read 1.

Step 4) Sequencing from the target fragment

Sequencing of the clusters from the above illustrative protocol was carried out using modified nucleotides prepared as described in International patent application WO 2004/018493, and labeled with four spectrally distinct fluorophores, as described in PCT application number PCT/GB2007/001770, published as WO07135368. Sequencing of clusters is described in more detail in patent WO06064199. The contents of the above-listed documents are incorporated herein by reference in their entireties.

A mutant 9°N polymerase enzyme (an exo- variant including the triple mutation L408Y/Y409A/P410V and C223S) (SBS polymerase) was used for the nucleotide incorporation steps.

All processes were conducted as described in the Illumina Genome Analyser operating manual. The flowcell was mounted to the analyser, primed with sequencing reagents: position #1 = incorporation mix (1 µM dNTP mix, 0.015 µg/ml SBS polymerase, 50 mM Tris pH 9.0, 50 mM NaCl, 6 mM MgSO4, 1 mM EDTA, 0.05 % Tween 20); position # 2 = spare (MilliQ water only); position # 3 = scan mix (100 mM Tris pH 7.0, 50 mM sodium acsorbate); position #4 = High salt wash (5x SSC, 0.05 % Tween 20); position # 5 = incorporation buffer (50 mM Tris pH 9.0, 50 mM NaCl, 1 mM EDTA, 0.05 % Tween 20); position #6 = cleavage mix (100 mM TCEP, 100 mM Tris pH 9.0, 100 mM NaCl, 50 mM sodium ascorbate, 0.05% Tween 20); position #7 = cleavage buffer (100 mM Tris pH 9.0, 100 mM NaCl, 0.05% Tween 20); position #8 = spare (single reads) or connected to PE module outlet (paired read experiments). Flowcells were sequenced using standard sequencing recipes for either 27- or 37-cycle experiments. Data was analysed using the standard analysis pipeline.

Incorporation: Prime with Incorporation buffer, 125 µL/channel; 60 µL/minutes, Heat to 60°C

Treat with Incorporation mix, 75 µL/channel; 60 µL/minutes.

Wait for a total of 15 minutes in addition to pumping fresh Incorporation mix, 25 µL/channel; 60 µL/minutes, every 4 minutes.

Cool to 20 °C.

Wash with Incorporation buffer, 75 µL/channel; 60 µL/minutes.

Wash with 5XSSC/0.05% Tween 20, 75 µL/channel; 60 µL/minutes

Prime with imaging buffer, 100 µL/channel; 60 µL/minutes

Scan in 4 colors at RT.

Cleavage: Prime with Cleavage buffer (0.1M Tris pH 7.4, 0.1M NaCl and 0.05% Tween 20), 125 µL/channel; 60 µL/minutes.

Heat to 60 °C.

Treat the clusters with Cleavage mix (100 mM TCEP in Cleavage buffer), 75 µL/channel; 60 µL/minutes.

Wait for a total of 15 minutes in addition to pumping fresh cleavage mix, 25 µL/channel; 60 µL/minutes, every 4 minutes.

Cool to 20 °C.

Wash with Enzymology buffer.

Wash with 5XSSC/0.05% Tween 20.

Repeat the process of Incorporation and Cleavage for as many cycles as required.

Incorporated nucleotides were detected using the Illumina genome analyzer, a Total Internal Reflection based fluorescent CCD imaging apparatus described in "Systems and Devices for Sequence by Synthesis Analysis," USSN 60/788,248, filed March 31, 2006 and corresponding PCT application PCT/US07/07991, published as WO07123744, the contents of which are incorporated herein by reference in their entirety.

Step 5) Sequencing of the Tag

Following the successful completion sequencing the first read on the Genome Analyser, flowcells remained mounted and were prepared for tag sequencing in situ, using the Illumina Paired End module. Temperature control was achieved by using the Genome Analyser peltier. All flow rates were 60 µl/min and 75 µl per lane unless otherwise stated. Clusters were denatured with 0.1 M NaOH to remove the extended sequencing primer from read 1. Hybridisation of the tag sequencing primer

(5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') (SEQ ID No. 62), 0.5 µM in hybridisation buffer) as described for read 1. Sequencing of the tag was performed on the Illumina Genome Analyser.

Step 6) Preparation of the clusters for the paired sample read

Following the successful completion sequencing of two reads on the Genome Analyser, flowcells remained mounted and were prepared for paired read 2 *in situ*, using the Illumina Paired End module. Temperature control was achieved by using the Genome Analyser peltier. All flow rates were 60 µl/min and 75 µl per lane unless otherwise stated. Clusters were denatured with 0.1 M NaOH to remove the extended sequencing primer from the tag sequencing. Clusters were 3'-dephosphorylated using T4 polynucleotide kinase (Invitrogen #18004-010, 200 U/ml, 50 mM imidazole pH 6.4, 12 mM MgCl2, 70 µM ADP, 1 mM 2-mercaptoethanol, 37 °C, 30 minutes), prior to re-synthesis of the A-strand achieved using 15 cycles of 60 °C isothermal amplification (same reagents and conditions as described for cluster creation except conducted at 30 µl/min). Clusters were washed before and after resynthesis with 0.3X SSC (150 µl and 245 µl respectively). Linearisation of the B-strand of the re-synthesised clusters was achieved by the excision of 8-oxoguanine from the B-type oligo using Fpg (formamidopyrimidine DNA glycosylase, NEB #M0240, 80 U/ml, 10 mM Bis Tris propane pH 7.0, 10 mM MgCl2, 1mM dithiothreitol, 37 °C, 30 minutes). Blocking was performed as described for read 1 using the same blocking cocktail. Linearised and blocked clusters were denatured prior to hybridisation of the read 2 specific sequencing primer

(5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') (SEQ ID No. 63), 0.5 µM in hybridisation buffer) as described for read 1.

Read 2 of the processed flowcells was subsequently sequenced on the Illumina Genome Analyser.

Data generated from the 12 samples is shown in figure 8. In each case the tag sequence could unambiguously determine the source of the sample. Each sample from the 12 was represented by a reasonable number of clusters, and for each tag read, the large majority (> 90%) of the reads aligned against the source genome identified by the tag.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above may be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.

Claims:

1. A method for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a universal adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by universal adaptor sequences and said universal adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid;

(c) amplifying adaptor-target-adaptors of each sample with two or more sample specific amplification primers to generate amplified nucleic acids, wherein one of said amplification primers comprises a sample specific tag sequence, and wherein amplified nucleic acids of each sample comprise said sample specific tag sequence and said sample specific tag sequence differentiates amplified nucleic acids originating from different samples;

(d) pooling the amplified nucleic acids of different samples;

(e) immobilising the amplified nucleic acids of different samples on an array to generate an array of immobilised fragments;

(f) sequencing the immobilised fragments on the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

2. A method for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples;

(c) pooling the adaptor-target-adaptors of different samples to generate a pooled sample of adaptor-target-adaptors;

(d) amplifying the pooled sample of adaptor-target-adaptors with a pair of universal amplification primers to generate an amplified pooled sample;

(e) immobilising the amplified pooled sample on an array to generate an array of immobilised fragments;

(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

3. A method for sequencing nucleic acid sequences on an array and identifying subsets of nucleic acid sequences on an array, wherein each subset of nucleic acid sequences is isolated from a different source, said method comprising the steps of:

(a) providing at least two samples of randomly fragmented double stranded nucleic acid targets, wherein each of said randomly fragmented double stranded nucleic acid targets is isolated from a different source;

(b) ligating a sample specific tagged adaptor to the ends of each target fragment of each sample to generate adaptor-target-adaptors of each sample, wherein each of said adaptor-target-adaptors comprises a target fragment flanked by tagged adaptor sequences and said tagged adaptor comprises a region of double stranded nucleic acid and at least one region of single stranded nucleic acid and a sample specific tag that differentiates adaptor-target-adaptors originating from different samples, and wherein the sample specific tag sequence is attached directly to the target fragment in each adaptor-target-adaptor with no intervening nucleotides between the sample specific tag sequence and target fragment sequences;

(c) amplifying adaptor-target-adaptors of each sample with a pair of amplification primers complementary to the sample specific tag sequence to generate amplified nucleic acids from each sample;

(d) pooling the amplified nucleic acids from each sample to generate a pool of amplified nucleic acid samples;

(e) immobilising the pool of amplified nucleic acid samples on an array to generate an array of immobilised fragments;

(f) sequencing the array to determine a sequence read of each immobilised target fragment and identify the sample specific tag sequence of each immobilised fragment, thereby determining both a nucleic acid sequence of said immobilized fragment and identifying the immobilized fragment as a member of a subset of nucleic acids on the array.

4. The method according to claim 1 or 2, wherein said sequencing is performed in two separate reads.

5. The method according to claim 3, wherein said sequencing of the target fragment and the sample specific tag sequence is performed in a single sequencing read.

6. The method according to claims 1 to 3, wherein the pooled samples are amplified on a solid surface prior to sequencing.

7. The method according to claim 6, wherein said solid surface is a collection of beads.

8. The method according to claim 7, wherein said beads are isolated as single beads in an emulsion.

9. The method according to claims 1 to 3, wherein the array of immobilized fragments is amplified prior to sequencing to generate an amplified array.

10. The method according to claim 9, wherein said amplified array is a clustered array of amplified single molecules.

11. The method according to claim 10, wherein the clustered array is formed by solid-phase nucleic acid amplification.

12. The method according to claim 11, wherein the clustered array is formed by isothermal solid-phase nucleic acid amplification.

13. The method according to any preceding claim, wherein said sequencing involves cycles of ligation with labelled oligonucleotides.

14. The method according to any preceding claim, wherein said sequencing involves cycles of addition of nucleotides.

15. The method according to claim 14, wherein said nucleotides are labelled.

16. The method according to claims 13 or 15, wherein said labels are fluorophores.

17. The method according to any preceding claim, wherein the sequencing step is carried out on an array of immobilised beads.

18. The method according to claim 1 wherein the nucleic acid target fragments can undergo a sequencing reaction from both ends to obtain a paired end read.

19. The method according to claim 18 wherein both ends of the nucleic acid fragment are sequenced before the sequence of the tag is determined.

20. The method according to claim 18 wherein the read from the second end of the fragment occurs after the tag is determined.

21. The method according to claim one that uses two amplification primers per sample, one of which carries the tag.

22. The method according to claim one that uses three amplification primers per sample, one of which carries the tag.

23. A kit comprising a double stranded nucleic acid adapter and at least two different primer polynucleotides that carry a different tag sequence that does not hybridise to the adapter.

24. A kit according to claim 23 further comprising a nucleotide polymerase and four nucleoside triphosphates.

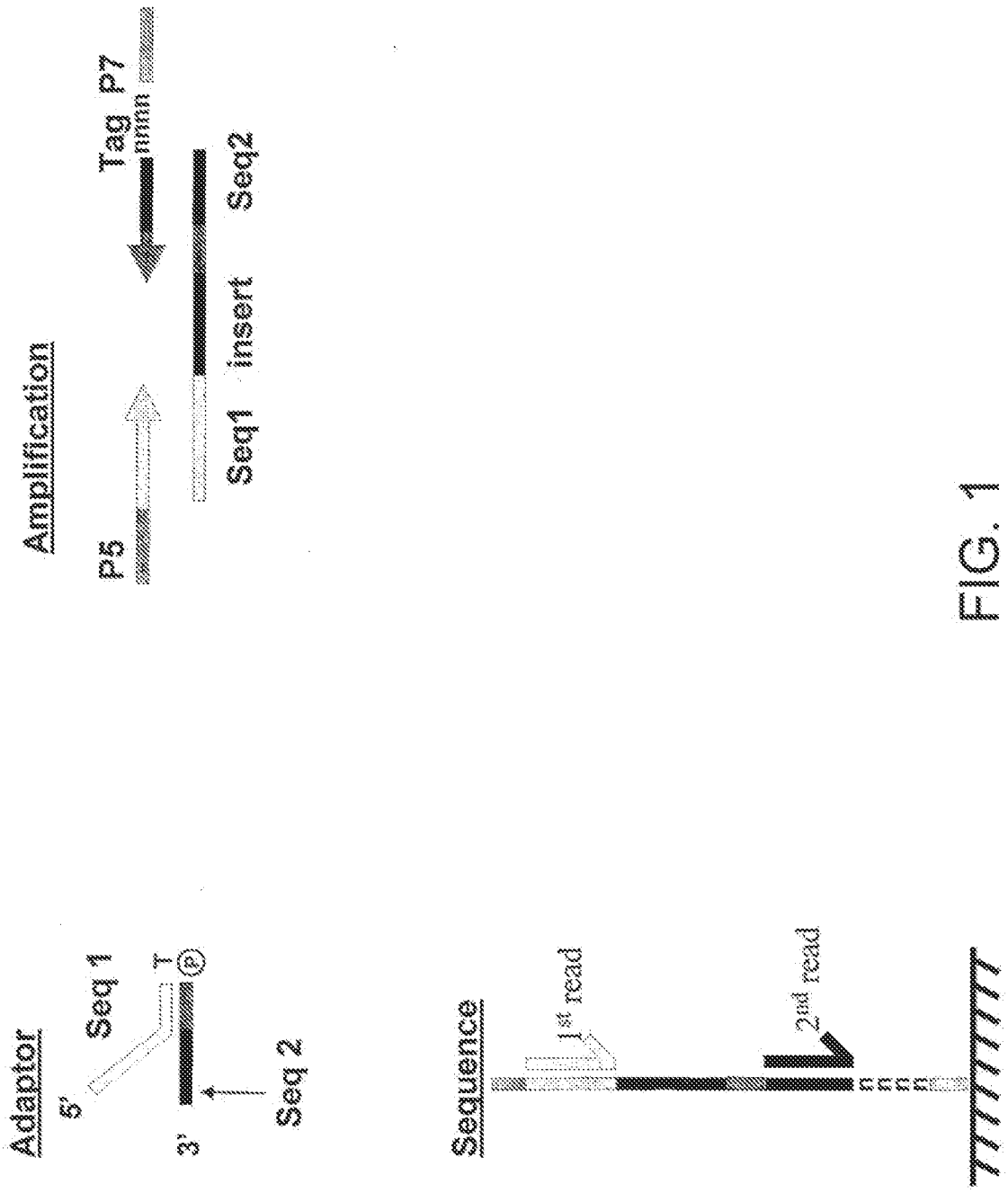


FIG. 1

IDT Exhibit 1018.0739

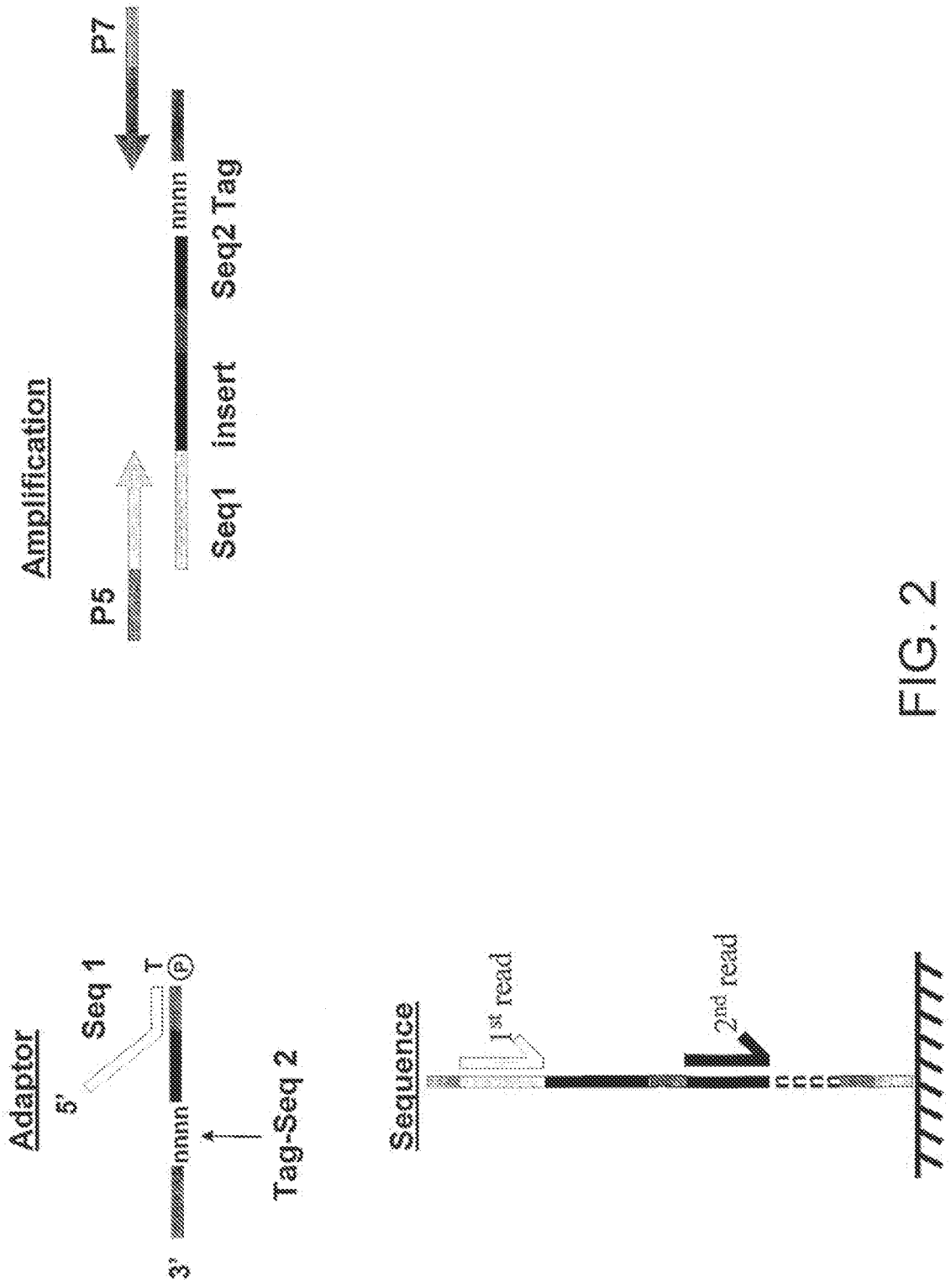


FIG. 2

IDT Exhibit 1018.0740



FIG. 3

IDT Exhibit 1018.0741

- Mixed two libraries and sequenced by SBS in one channel
- Data from a single tile

Reference Sequences	% of Clusters Aligning to Reference
PhiX	68
BAC	29
E.coli	3

	PhiX TAG	BAC TAG	Other TAGs
PhiX	98 %	0.1 %	2 %
BAC	0.1 %	98 %	2 %

- Majority of BAC and PhiX inserts were sequenced with the correct TAG

FIG. 4

IDT Exhibit 1018.0742

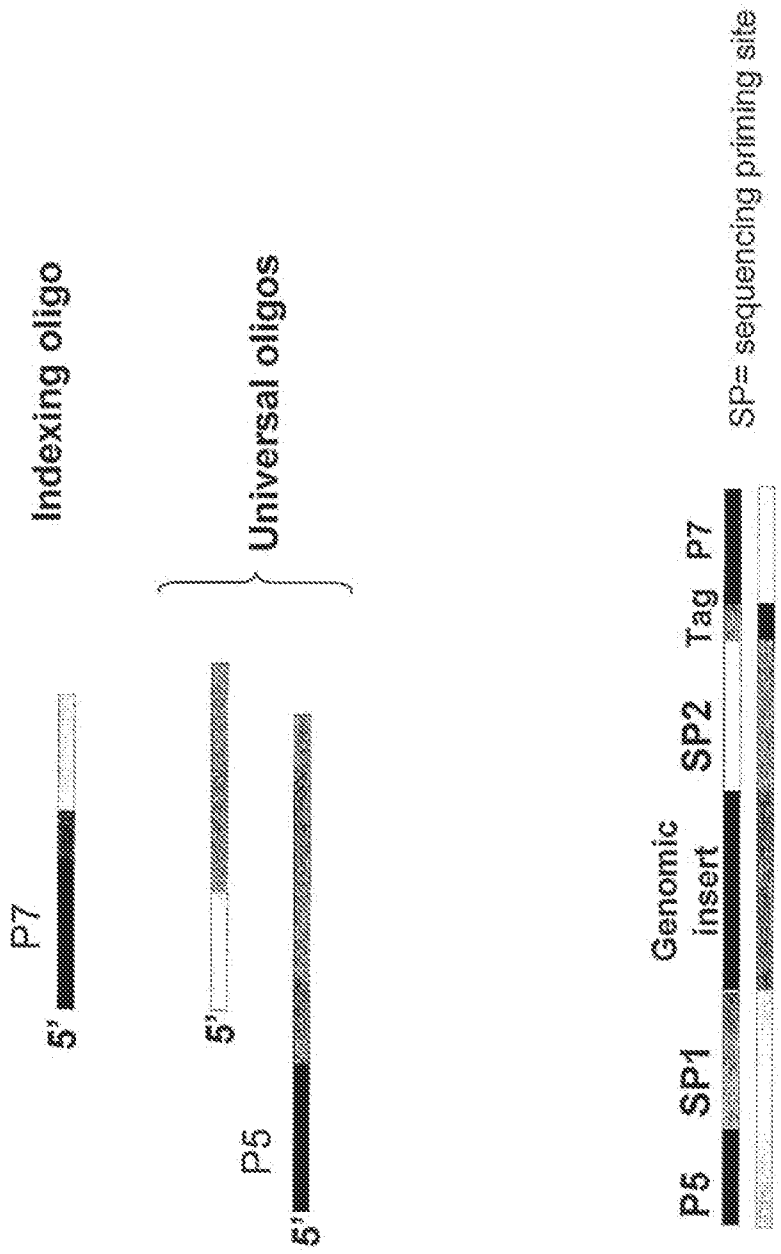


FIG. 5

IDT Exhibit 1018.0743

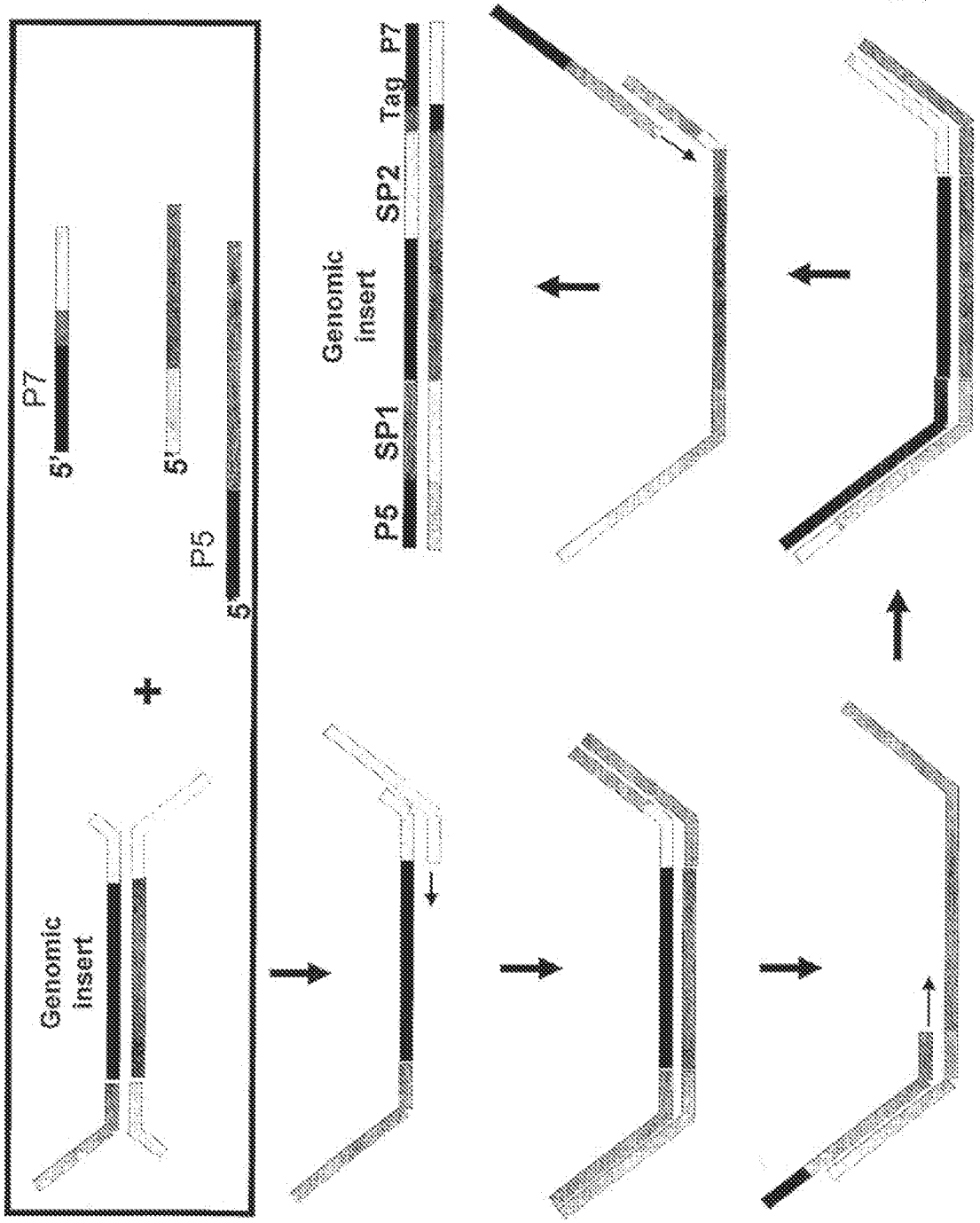


FIG. 6a

IDT Exhibit 1018.0744

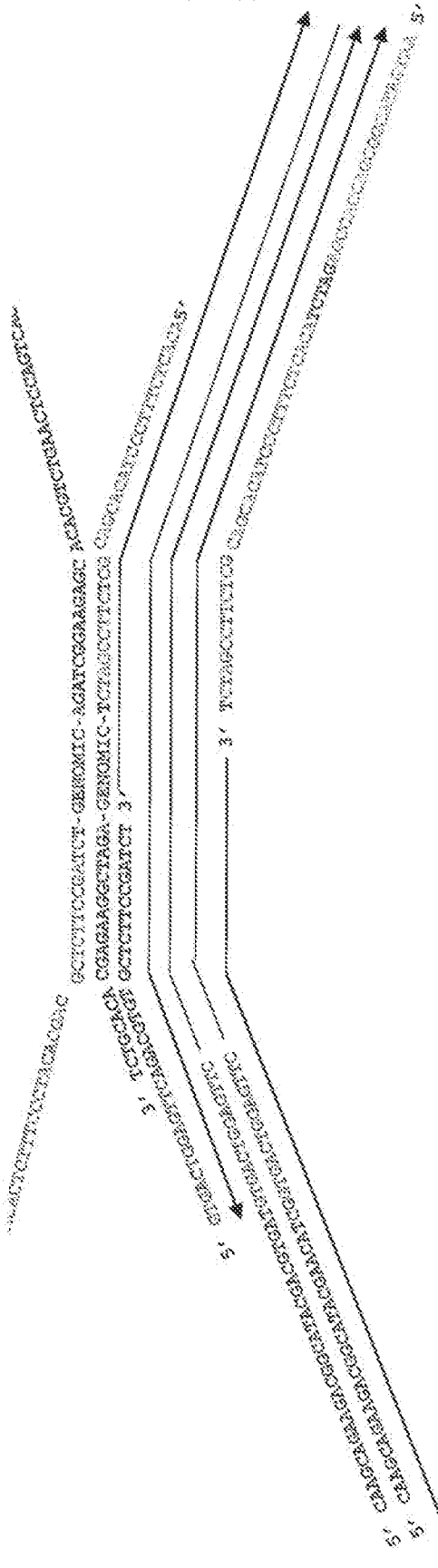


FIG. 6b

IDT Exhibit 1018.0745

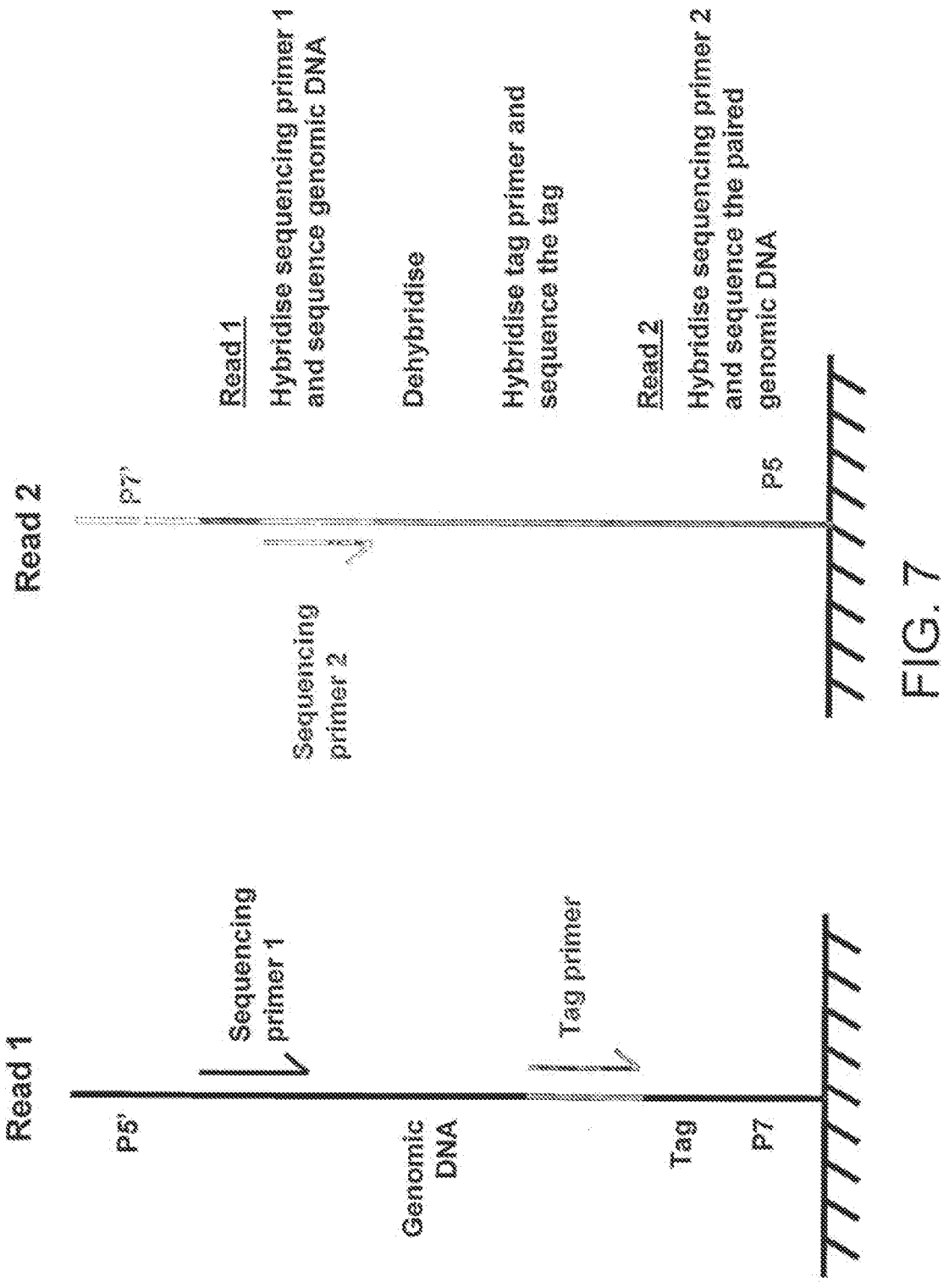


FIG. 7

IDT Exhibit 1018.0746

Tag	Genome	Clus No	BAC R1	BAC R2	PhiX R1	PhiX R2
ATCACG	PhiX	6522	0%	0%	96%	92%
CGATGT	PhiX	3548	0%	0%	96%	92%
TTAGGC	PhiX	10394	0%	0%	97%	93%
TGACCA	PhiX	6842	0%	0%	97%	93%
ACAGTG	PhiX	6677	0%	0%	97%	93%
CAGATC	PhiX	11231	0%	0%	96%	93%
GCCAAT	BAC	6174	97%	92%	0%	0%
ACTTGA	BAC	5375	97%	93%	0%	0%
GATCAG	BAC	5404	97%	93%	0%	0%
TAGCTT	BAC	10239	97%	94%	0%	0%
GGCTAC	BAC	9734	97%	93%	0%	0%
CTTGTA	BAC	9704	97%	93%	0%	0%

Expected Tag	Observed Tag	Genome	Clus No	BAC R1	BAC R2	PhiX R1	PhiX R2
TGACCA	TGACAA	PhiX	8	0%	0%	100%	88%
GCCAAT	TCCAAT	BAC	34	97%	94%	0%	0%
GATCAG	TATCAG	BAC	28	100%	96%	0%	0%

FIG. 8

IDT Exhibit 1018.0747

Forked Adaptor Primers SBS3/SBS12 (indexing/paired ends)

Adaptor oligo 2 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3'

Adaptor oligo 1 (short) 5' P-GATCGGAAGAGCACACGTCT 3'

5' ACACTCTTTCCCTACACGAC GCTCTTCCGATCT 3'

3' TCTGCACACGAGAAGGCTAG 5'

PCR Primer		
Universal LPX primer	5'	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCxT 3'
Lpx primer	5'	xATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT
Tag1 primer (CGTGAT)	5'	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC 3'
Tag2 primer (ACATCG)	5'	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC 3'
Tag3 primer (GCCTAA)	5'	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTC 3'
Tag4 primer (TGGTCA)	5'	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC 3'
Tag5 primer (CACTGT)	5'	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTC 3'
Tag6 primer (ATTGGC)	5'	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC 3'
Tag7 primer (GATCTG)	5'	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTC 3'
Tag8 primer (TCAAGT)	5'	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC 3'
Tag9 primer (CTGATC)	5'	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTC 3'
Tag10 primer (AAGCTA)	5'	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTC 3'
Tag11 primer (GTAGCC)	5'	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTC 3'
Tag12 primer (TACAAG)	5'	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTC 3'

Sequencing primers		
SBS12	5'	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3'
SBS12rev	5'	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC 3'
SBS3	5'	ACACTCTTTCCCTACACGACGCTCTTCCGATC 3'

FIG. 9

IDT Exhibit 1018.0748

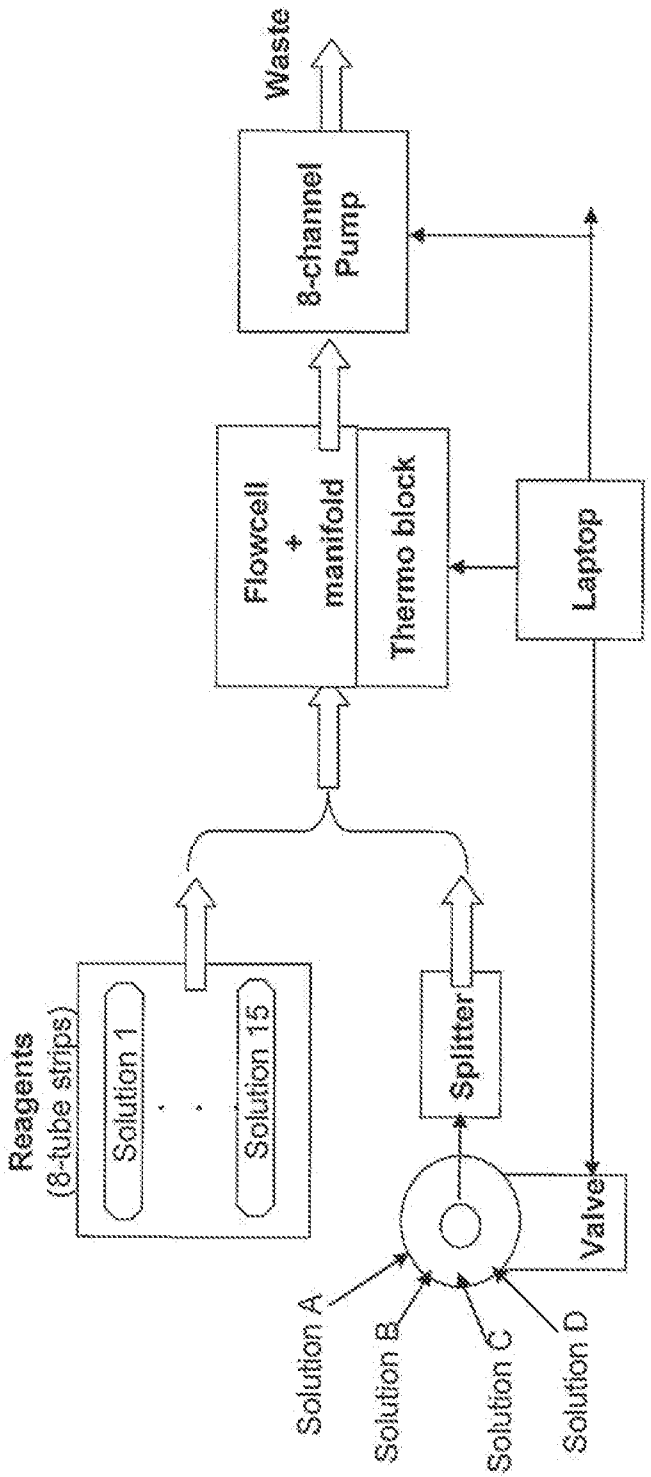


FIG. 10

IDT Exhibit 1018.0749

Electronic Patent Application Fee Transmittal

Application Number:	14540917
Filing Date:	13-Nov-2014
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Filer:	Jeffrey R. Bolland
Attorney Docket Number:	NUGN-001/01US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**180**

Electronic Acknowledgement Receipt

EFS ID:	23247867
Application Number:	14540917
International Application Number:	
Confirmation Number:	7076
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Customer Number:	58249
Filer:	Jeffrey R. Bolland
Filer Authorized By:	
Attorney Docket Number:	NUGN-001/01US
Receipt Date:	19-AUG-2015
Filing Date:	13-NOV-2014
Time Stamp:	13:00:37
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	$180
RAM confirmation Number	9414
Deposit Account	501283
Authorized User	COOLEY LLP
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		NUGN-001_01US_Response_to_NFOA.pdf	142680 06d7c5bc36f60c5c36d831ff70f6a6be2af32cfe	yes	11

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Claims	2	5
Applicant Arguments/Remarks Made in an Amendment	6	11

Warnings:

Information:

2	Information Disclosure Statement (IDS) Form (SB08)	NUGN-001_01US_IDS_August_2015.pdf	69742 2fd9cb51a7792dde42a6323b9a8f7f2464642d3c	no	4

Warnings:

Information:

This is not an USPTO supplied IDS fillable form

3	Non Patent Literature	Myers_Hudon_Alpha_2013.pdf	3679773 86c0dd39853c16b157dd64e7f5757fe376d744d4	no	15

Warnings:

Information:

4	Foreign Reference	WO2014144092-PAMPH-164_.pdf	10419986 e6684bdc4b929d13bd60222f106a3e1be7954fb7	no	68

Warnings:

Information:

5	Foreign Reference	WO2012061832-PAMPH-249_.pdf	11652152 828cdb94517cf80709698e82478825c0075afbcd	no	103

Warnings:

Information:

6	Foreign Reference	WO2008093098-PAMPH-183_.pdf	13606793 79b481541cfac705dde070df4aec986c1f1e32c6	no	111
Warnings:					
Information:					
7	Non Patent Literature	NUGN-001_01WO_ISR_WO.pdf	2767104 8d3a06bfffd70945b91edc36d24d78e690d08045	no	17
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30795 e616816b9b263bc57dc608a80ff285c2ecc8396a	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	42369025	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20120208705	A1	2012-08-16	STEEMERS et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2008093098	WO	A1	2008-08-07	SOLEXA LIMITED		☐
	2	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		☐
	3	2014144092	WO	A1	2014-09-18	NUGEN TECHNOLOGIES, INC.		☐

IDT Exhibit 1018.0755

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	MYERS, R. M. Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology, May 2, 2013. p. 1-15. Retrieved from ,https://www.encodeproject.org/documents/9fb8e06c-d145-447a-a606-e7f3dc44b549/@@download/attachment/HA_Myers_miRNA-seq_Protocol_2013-05-02.pdf> on February 28, 2015.	☐
	2	International Searching Authority. Written Opinion and International Search Report for PCT/US2014/065530. Date of mailing is March 5, 2015.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

IDT Exhibit 1018.0756

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Jeffrey R. Bolland/	Date (YYYY-MM-DD)	2015-08-19
Name/Print	Jeffrey R. Bolland	Registration Number	73,759

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

IDT Exhibit 1018.0757

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IDT Exhibit 1018.0758

IN THE CLAIMS:

1. (Currently Amended) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

    a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

        (i) an indexing primer binding site;

        (ii) an indexing site;

        (iii) an identifier site comprising at least 8 nucleotides; and

        (iv) a target sequence primer binding site;

    b) amplifying the adapter-nucleic acid fragment ligated products;

    c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

    d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence;

    wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is not conserved across all adaptor sequences.

2. (Original) The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.

3. (Original) The method of claim 1, wherein the identifier site is sequenced with the indexing site.

4. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the indexing site.

5. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced with the target sequence.

6. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.

7. (Original) The method of claim 1, wherein the adaptor comprises from 5' to 3':

    (i) the indexing primer binding site;

    (ii) the indexing site;

    (iii) the identifier site; and

    (iv) the target sequence primer binding site.

8. (Withdrawn) The method of claim 1, wherein the adaptor comprises from 5' to 3':

    (i) the indexing primer binding site;

    (ii) the indexing site;

    (iii) the target sequence primer binding site; and

    (iv) the identifier site.

9. (Original) The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.

10. (Original) The method of claim 9, wherein the nucleic acid fragments from each sample has the same indexing site.

11. (Original) The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. (Original) The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. (Original) The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. (Original) The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. (Original) The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. (Original) The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. (Currently Amended) The method of claim 1, wherein the identifier site is ~~between 1 and 8~~ at least 10 nucleotides in length.

18. (Original) The method of claim 1, wherein the identifier site is about 8 nucleotides in length.

19. (Original) The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. (Original) The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

21. (Currently Amended) A kit comprising a plurality of double-stranded adaptors, wherein each adaptor comprises:

    (i) an indexing primer binding site;

    (ii) an indexing site;

    (iii) an identifier site comprising at least 8 nucleotides; and

(iv) a target sequencing primer binding site;

wherein the kit further comprises one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained; wherein the indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is not conserved across all adaptor sequences.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:	AMORESE, Douglas, et al.	Confirmation No.:	7076
Application No.:	14/540,917	Group Art Unit:	1637
Filed:	November 13, 2014	Examiner:	Teresa E. Strzelecka

FOR: **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT/RESPONSE TO OFFICE ACTION

In response to the Office Action dated May 20, 2015, this paper is timely filed within the 3-month period for reply ending August 20, 2015. Please amend the above-identified patent application in the following manner:

Amendments to the Claims are reflected in the listing of the claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

REMARKS

I. Status

Claims 1-21 are pending; claims 4-6 and 8 have been withdrawn from consideration; and claims 1-3, 7, and 9-21 have been examined on the merits.

Claims 1, 17, and 21 have been amended. Support for the amendment of claim 1 can be found in the specification in at least paragraphs [0029], [0030], and paragraphs [0084] – [0091] which describe the random or semi-random nature of each of the identifier sequences/sites. Support for the amendment of claim 17 can be found in at least paragraph [0030]. Support for the amendment of claim 21 can be found in the specification in at least paragraphs [0028] – [0030], [0071], [0072], [0074], and [0084] – [0091].

No new matter is introduced by way of amendment. Accordingly, entry thereof is respectfully requested. Upon entry of this amendment, claims 1-3, 7, and 9-21 are under examination.

II. Restriction/Election Requirement

The Examiner is thanked for the rejoinder and consideration of claim 21, which is directed to a kit comprising adaptors of the presently claimed subject matter.

III. 35 U.S.C. § 101

Claim 21 is rejected as allegedly directed to a judicial exception. The Examiner asserts that the claim is drawn to a "plurality of adaptors, i.e., nucleic acid sequences, comprising an indexing primer binding site, an indexing site, an identifier site and a target sequencing binding site." Office Action, p. 4. The Examiner then alleges that because no specific sequences are

claimed, "the claim is drawn to a plurality of oligonucleotides which have corresponding sequences occurring in nature." *Id.* The Examiner also states that the claim does not "include additional elements that are sufficient to amount to significantly more than the judicial exception because no other structural elements are claimed." Office Action, p. 5. Applicant respectfully traverses this rejection.

Solely for expediting prosecution, claim 21 has been amended to further recite oligonucleotide primers, reagents for amplification, instructions, and the limitations of the plurality of adaptors of instant claim 1. Instant claim 21 is drawn to a unique grouping of specific oligonucleotides. The plurality of adaptors claimed in the kit comprise an indexing primer binding site, an indexing site, an identifier site comprising at least 8 nucleotides, and a target sequencing primer binding site. The indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides, and the sequence of the identifier site is not conserved across all adaptor sequences. The instantly claimed plurality of adaptors cannot be equated to corresponding sequences occurring in nature. Furthermore, the amended claim now reads on a kit that cannot simply be interpreted as any composition of oligonucleotides. Accordingly, withdrawal of the rejection is respectfully requested.

V. 35 U.S.C. § 102(a)(1)

The Examiner has rejected claim 21 as allegedly being anticipated by both Grothues *et al.* (1993. *Nucl. Acids Res.* 21:1321-1322) and McCloskey *et al.* (2007. *Biochem. Genet.* 45:761-767). Applicant respectfully traverses this rejection.

The Examiner states Grothues *et al.* discloses a plurality of tagged random primers, comprising 9-15 bp random 3' component and a 17 bp constant sequences at the 5' end, and

alleges these primers are the same as the adaptors as claimed, and thus, anticipates the instant claims. Office Action, p. 5. Applicant respectfully disagrees. The instant claims are not anticipated by Grothues *et al.* as the primers disclosed by Grothues et al. are single stranded and the adaptors of the instant claims are double-stranded. In addition, Grothues et al. does not disclose, teach, or suggest instructions for employing the kit; separately contained adaptors, reagents for amplification, and oligonucleotide primers; the indexing site as unique amongst only a subset of the adaptors and as an index for multiple polynucleotides; and the identifier site as not being conserved across all adaptor sequences. Accordingly, Grothues *et al.* fails to teach each of the elements of instant claim 21 and does not anticipate claim 21.

The Examiner states McCloskey *et al.* discloses a plurality of randomly-barcoded oligonucleotides containing, from the 3' end, a target-specific sequence, batch-stamp sequence, a random 7 bp barcode, another primer binding site, and another sequence, and alleges these oligonucleotides are the adaptors as claimed, and thus, anticipates the instant claims. Office Action, p. 6. Applicant respectfully disagrees. The instant claims are not anticipated by McCloskey *et al.* as McCloskey *et al.* does not disclose, teach, or suggest instructions for employing the kit; separately contained adaptors, reagents for amplification, and oligonucleotide primers; the indexing side as unique amongst only a subset of the adaptors and as an index for multiple polynucleotides; and the identifier site as comprising at least 8 nucleotides. Accordingly, McCloskey *et al.* fails to teach each of the elements of instant claim 21 and does not anticipate claim 21.

Withdrawal of the rejection is respectfully requested.

VI. 35 U.S.C. § 103

The Examiner has rejected claim 1-3, 7, and 9-20 as allegedly being unpatentable over Samuels *et al.* (US20120220494A1, filed 02/16/12 and published 08/30/12) and McCloskey *et al.* Applicant respectfully traverses this rejection.

The Examiner states "Samuels et al. teach a method for detecting a duplicate sequencing read from a population of sample sequence reads." Office Action at p. 6. However, the Examiner also states "Samuels et al. teach detection of PCR-introduced errors in sequencing, ***but do not specifically teach detecting duplicate sequencing reads***." Office Action, p. 8 (emphasis added). Samuels *et al.* discloses a method of providing barcode sequences to targets such as a nucleic acid sequences, wherein universal priming sites may be present in barcode libraries [205], [0216]-[0216]. The barcode sequences and universal priming sites may be attached to the nucleic acid sequences and amplified in a PCR reaction [0288]-[0289]. The amplification products may then be sequenced as a whole or in part [0326-0327]. Sequencing errors versus single nucleotide polymorphisms may be detected through the use of barcodes and sequencing of the amplification products [0339] – [0341].

Samuels *et al.* do not teach (1) the inclusion of an identifier site comprising at least 8 nucleotides, wherein the identifier site is not conserved across all adaptor sequences; (2) the indexing site as unique amongst only a subset of the nucleic acid fragments; (3) a target sequence primer binding site; and (3) the detection and removal of duplicate sequencing reads.

McCloskey *et al.* disclose a method of identifying and removing duplicate reads through incorporating an adaptor comprising a batch stamp, a barcode, a primer-binding site for PCR, and extra bases that serve to minimize nonspecific annealing, wherein the barcodes comprise 7 random nucleotides (p. 763, Fig. 1; p. 764, para 2; p. 766, para 4-5).

McCloskey *et al.* does not teach the method wherein the identifier site (barcode) comprises at least 8 nucleotides or a target sequence primer binding site.

The Examiner alleges McCloskey *et al.* accounts for the deficiencies of Samuels *et al.* Applicant respectfully disagrees. Samuels *et al.* and McCloskey *et al.*, alone or in combination fail to teach an indexing site that is at least 8 nucleotides in length, as recited in the instant claims. McCloskey *et al.* only discloses barcodes that are 7 nucleotides in length, which limit the total number of adaptor sequences comprising unique barcodes that can be generated. The instant claims recite indexing sites of at least 8 nucleotides in length, which allows for a significantly greater population of adaptor sequences comprising unique barcodes that can be generated.

Withdrawal of the rejection is respectfully requested.

CONCLUSION

In view of the foregoing, Applicant respectfully submits that no further impediments exist to the allowance of this application and, therefore, requests an indication of allowability. However, the Examiner is requested to call the undersigned if any questions or comments arise.

The Director is hereby authorized to charge any appropriate fees under 37 C.F.R. §§1.16, 1.17, and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 50-1283.

Dated: August 19, 2015

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (202) 728-7017
Fax: (202) 842-7899

By: /Jeffrey R. Bolland/
Jeffrey R. Bolland, Ph.D.
Reg. No. 73,759

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/540,917	Filing Date 11/13/2014	☐ To be Mailed

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT **08/19/2015**

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 21	Minus	** 21	= 0	X $80 =	0
Independent (37 CFR 1.16(h))	* 2	Minus	***3	= 0	X $420 =	0
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	0

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
/ANTJUAN RIVERA/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /TS/

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20120208705	A1	2012-08-16	STEEMERS et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2008093098	WO	A1	2008-08-07	SOLEXA LIMITED		☐
	2	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		☐
	3	2014144092	WO	A1	2014-09-18	NUGEN TECHNOLOGIES, INC.		☐

IDT Exhibit 1018.0771

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14540917
Filing Date	2014-11-13
First Named Inventor	AMORESE, Douglas
Art Unit	1637
Examiner Name	Teresa E. Strzelecka
Attorney Docket Number	NUGN-001/01US 301735-2003

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /TS/

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	MYERS, R. M. Myers lab microRNA-seq Protocol, Hudson Alpha Institute for Biotechnology, May 2, 2013. p. 1-15. Retrieved from ,https://www.encodeproject.org/documents/9fb8e06c-d145-447a-a606-e7f3dc44b549/@@download/attachment/HA_Myers_miRNA-seq_Protocol_2013-05-02.pdf> on February 28, 2015.	☐
	2	International Searching Authority. Written Opinion and International Search Report for PCT/US2014/065530. Date of mailing is March 5, 2015.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/Teresa Strzelecka/	Date Considered	09/21/2015

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

IDT Exhibit 1018.0772

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/540,917	11/13/2014	Douglas AMORESE	NUGN-001/01US	7076

58249 7590 09/28/2015

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue, NW
Suite 700
Washington, DC 20004

EXAMINER
STRZELECKA, TERESA E

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
09/28/2015	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zpatdcdocketing@cooley.com

IDT Exhibit 1018.0773

Office Action Summary	Application No. 14/540,917	Applicant(s) AMORESE ET AL.	
	Examiner TERESA E. STRZELECKA	Art Unit 1637	AIA (First Inventor to File) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *8/19/15*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-21* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-21* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date *8/19/15*.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

IDT Exhibit 1018.0774

1. The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

DETAILED ACTION

2. This office action is in response to an amendment filed August 19, 2015. Claims 1-21 were previously pending, with claims 4-6 and 8 withdrawn from consideration. Applicant amended claims 1, 17 and 21. Claims 1-3, 7 and 9-21 are under consideration.

3. Applicant's amendments overcame the rejection of claim 21 under 35 U.S.C. 102(a)(1) as being anticipated by Grothues et al. and the rejection of claim 21 under 35 U.S.C. 102(a)(1) as being anticipated by McCloskey et al. All other previously presented rejections are maintained for reasons given in the "Response to Arguments" section below.

4. This office action contains new grounds for rejection necessitated by amendment.

Information Disclosure Statement

5. The information disclosure statement (IDS) submitted on August 19, 2015 was filed after the mailing date of the non-final office action on May 20, 2015. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Response to Arguments

6. Applicant's arguments filed August 19, 2015 have been fully considered but they are not persuasive.

A) Regarding the rejection of claim 21 under 35 U.S.C. 101 Applicant argues the following:

"Solely for expediting prosecution, claim 21 has been amended to further recite oligonucleotide primers, reagents for amplification, instructions, and the limitations of the plurality of adaptors of instant claim 1. Instant claim 21 is drawn to a unique grouping of specific oligonucleotides. The plurality of adaptors claimed in the kit comprise an indexing primer binding site, an indexing site, an identifier site comprising at least 8 nucleotides, and a target sequencing primer binding site. The indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides, and the sequence of the identifier site is not conserved across all adaptor sequences. The instantly claimed plurality of adaptors cannot be equated to corresponding sequences occurring in nature. Furthermore, the amended claim now reads on a kit that cannot simply be interpreted as any composition of oligonucleotides. Accordingly, withdrawal of the rejection is respectfully requested."

However, all of these other added elements are products on nature themselves (such as primers) or are elements which are well-understood, routine and conventional elements used in the kits for methods of nucleic acid amplification and sequencing (see Bignell et al., US 2009/0233802A1, cited below). Therefore, the new limitations do not amount to "significantly more" than the judicial exceptions of adapters and primers.

The rejection is maintained.

B) Regarding the rejection of claims 1-3, 7 and 9-20 under 35 U.S.C. 103 as being unpatentable over Samuels et al. and McCloskey et al., Applicant argues the following:

"Samuels *et al.* do not teach (1) the inclusion of an identifier site comprising at least 8 nucleotides, wherein the identifier site is not conserved across all adaptor sequences; (2) the indexing site as unique amongst only a subset of the nucleic acid fragments; (3) a target sequence primer binding site; and (3) the detection and removal of duplicate sequencing reads.

McCloskey *et al.* disclose a method of identifying and removing duplicate reads through incorporating an adaptor comprising a batch stamp, a barcode, a primer-binding site for PCR, and extra bases that serve to minimize nonspecific annealing, wherein the barcodes comprise 7 random nucleotides (p. 763, Fig. 1; p. 764, para 2; p. 766, para 4-5).

McCloskey *et al.* does not teach the method wherein the identifier site (barcode) comprises at least 8 nucleotides or a target sequence primer binding site.

The Examiner alleges McCloskey *et al.* accounts for the deficiencies of Samuels *et al.* Applicant respectfully disagrees. Samuels *et al.* and McCloskey *et al.,* alone or in combination fail to teach an indexing site that is at least 8 nucleotides in length, as recited in the instant claims. McCloskey *et al.* only discloses barcodes that are 7 nucleotides in length, which limit the total number of adaptor sequences comprising unique barcodes that can be generated. The instant claims recite indexing sites of at least 8 nucleotides in length, which allows for a significantly greater population of adaptor sequences comprising unique barcodes that can be generated."

However, Samuels et al. specifically teach unique N-mers which contain barcodes 2-21 nucleotides in length (page 3, [0029]. Samuels et al. also specifically teach that barcodes contain 4-20 nucleotides which uniquely encode library members (page 15, [0209]) and further

specifically teach that the number of unique barcodes depends on the number of bases in a barcode, with 6, 7, 8 10 or 15 nucleotides resulting in unique identification of 4096, 16384, 65356, about one million and over than one billion libraries (page 15/16, [0217]). Therefore, since McCloskey et al. teach that barcodes can be used to filter out duplicate reads, and Samuels et al. teach the required lengths of barcodes, the rejection is maintained.

Claim Rejections - 35 USC § 112

7. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

8. Claim 21 is rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor or a joint inventor, or for pre-AIA the inventor(s), at the time the application was filed, had possession of the claimed invention.

Claim 21 was amended to read "A kit comprising a plurality of double-stranded adaptors…". However, there is no support in the specification or in the originally filed claims for the limitation of an adapter being double-stranded, therefore it represents new matter.

Claim Rejections - 35 USC § 101

9. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

10. Claim 21 is rejected under 35 U.S.C. 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more. Claim(s) 21 is directed to a kit comprising a plurality of adaptors, i.e., nucleic acid sequences, comprising an indexing primer binding site, an indexing site, an identifier site and a target sequencing binding site, the kit further comprising one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained.

No specific sequences of the adapters and oligonucleotide primers are claimed; therefore the claim is drawn to a plurality of oligonucleotides which have corresponding sequences occurring in nature. The claim(s) does/do not include additional elements that are sufficient to amount to significantly more than the judicial exception because the other elements of the kit are well-understood, routine and conventional in the art. Specifically, "reagents for amplification" are not specified, therefore could encompass water and salt solutions, for example. It is conventional in the art to include instructions for using a kit, as well as to partition components of a kit into separate containers (see, for example, US 2009/0233802 A1, [0103]-[0107]; US 5,384,242 A; issued January 1995; col. 11, lines 61-68; col. 12, lines 1-13). Therefore, the

additional elements recited in claim 21 do not amount to "significantly more" than the claimed judicial exceptions of nucleic acid sequences of adapters and primers.

In conclusion, claim 21 is not patent eligible under 35 U.S.C. 101.

Claim Rejections - 35 USC § 103

11. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

12. The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102 of this title, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

13. Claims 1-3, 7 and 9-20 are rejected under 35 U.S.C. 103 as being unpatentable over Samuels et al. (US 2012/0220494 A1; filed February 16, 2012 and published August 30, 2012; previously cited) and McCloskey et al. (Biochem. Genet., vol. 45, pp. 761-767, 2007; previously cited).

A) Regarding claims 1 and 7, Samuels et al. teach a method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

(i) an indexing primer binding site;

(ii) an indexing site;

(iii) an identifier site comprising at least 8 nucleotides; and

(iv) a target sequence primer binding site (page 14, [0205]; page 15, [0209], [0212], [0216]-[0217]; Fig. 6; page 16, [0231]-[0232]; page 17, [0232]-[0235]; page 16/17, [0245]; where the adaptors have an indexing primer binding site, an indexing site, an identifier site and a target sequence primer binding site; Samuels et al. teach identifier site between 2 and 21 nucleotides in length, or from 5-15 nucleotides (page 3, [0029]; page 15/16, [0217]). The indexing site and identifiers are both unique to subsets of polynucleotides are not conserved among all adaptor sequences, since they constitute random nucleotides);

b) amplifying the adapter-nucleic acid fragment ligated products (page 22, [0288]-[0289]; page 25, [0326]-[0327]);

c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products (page 25, [0326]-[0328]; page 26, [0328]-[0337]; and

d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence (page 26/27, [0328]; page 27, [0339]-[0341]; Fig. 43 and 44; where sequencing errors due to PCR are detected).

Regarding claim 3, Samuels et al. teach sequencing the identifier site with the indexing site (Fig. 43; page 27, [0339]-[0341]).

Regarding claim 9, Samuels et al. teach barcoding multiple samples (page 6, [0122]; page 14, [0202]).

Regarding claims 13 and 14, Samuels et al. teach DNA and RNA fragments, and genomic DNA (page 2, [0017]).

Regarding claims 17 and 18, Samuels et al. teach identifier site between 2 and 21 nucleotides in length, or from 5-15 nucleotides (page 3, [0029]; page 15/16, [0217]).

Regarding claim 19, Samuels et al. teach indexing primer site being a universal primer site (page 15, [0216]).

Regarding claim 20, Samuels et al. teach target sequence binding site being a universal binding site (page 15, [0216]).

B) Samuels et al. teach detection of PCR-introduced errors in sequencing, but do not specifically teach detecting duplicate sequencing reads.

C) Regarding claim 1, McCloskey et al. teach attaching an adaptor containing an indexing primer binding site, an indexing site, an identifier site and a target sequence binding site to genomic DNA by amplification, PCR-amplifying the resulting construct and sequencing (Fig. 1; page 762, paragraphs 3-5; page 763; page 764, paragraphs 1-4).

Regarding claims 1, 2 and 10-12, McCloskey et al. teach sequencing the PCR products containing the adapter sequences, grouping the sequencing reads based on the batch stamp and barcode and detecting and removing duplicate reads (page 766, fourth and fifth paragraphs).

Regarding claims 15 and 16, McCloskey et al. teach indexing site between six and eight bp (page 764, second paragraph).

It would have been *prima facie* obvious to one of ordinary skill in the art at the time of the invention to have used the batch-stamp and barcode-encoded oligonucleotides as suggested by McCloskey et al. in the method of Samuels et al. to eliminate PCR-originating artefacts. As stated by McCloskey et al. (Abstract):

"Polymerase chain reaction (PCR) has become the mainstay of DNA sequence analysis. Yet there is always uncertainty concerning the source of the template DNA that gave rise to a particular PCR product. The risks of contamination, biased amplification, and product redundancy are especially high when limited amounts of template DNA are used. We have developed and applied molecular encoding principles to solve this source-uncertainty problem for DNA sequences generated by standard PCR. Batch-stamps specify the date and sample identity, and barcodes detect template redundancy. Our approach thus enables classification of each PCR-derived sequence as valid, contaminant, or redundant, and provides a measure of sequence diversity. We recommend that batch-stamps and barcodes be used when amplifying irreplaceable DNAs and cDNAs available for forensic, clinical, single cell, and ancient DNA analyses."

14. Claim 21 is rejected under 35 U.S.C. 103 as being unpatentable over Samuels et al. (US 2012/0220494 A1; filed February 16, 2012 and published August 30, 2012; previously cited) and McCloskey et al. (Biochem. Genet., vol. 45, pp. 761-767, 2007; previously cited), Bignell et al. (US 2009/0233802 A1; published September 2009), Oakes (US 5,384,242 A; issued January 2005) and Stratagene Catalog (page 39, 1988).

A) Samuels et al. and McCloskey et al. teach the double-stranded adapters, but do not teach kits or the additional components recited in claim 21.

B) Regarding claim 21, Bignell et al. teach a kit comprising adapters, amplification primers, amplification reagents and sequencing primers (page 11, paragraphs [0104]-[0107]). Oakes teaches kits comprising primers, probes and amplification reagents in separate containers, as well as intructions for use of the kit.

Therefore it would have been *prima facie* obvious to one of ordinary skill in the art at the time of the invention to have packaged the components necessary for the method of Samuels et al. and McCloskey et al. into a kit as suggested by Bignell et al. and Oakes. The motivation for using kits is provided by Stratagene Catalog:

"Each kit provides two services: 1) a variety of different reagents have been assembled and pre-mixed specifically for a defined set of experiments. Thus one need not purchase gram quantities of 10 different reagents, each of which is needed in only microgram amounts, when beginning a series of experiments. When one considers all of the unused chemicals that typically accumulate in weighing rooms, desiccators, and freezers, one quickly realizes that it is actually far more expensive for a small number of users to prepare most buffer solutions from the basic reagents. Stratagene provides only the quantities you will actually need, premixed and tested. In actuality, the kit format saves money and resources for everyone by dramatically reducing waste. 2) The other service provided in a kit is quality control" (page 39, column 1).

15. No claims are allowed.

Conclusion

16. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period

will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to TERESA E. STRZELECKA whose telephone number is (571)272-0789. The examiner can normally be reached on M-F (8:30-5:30).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on (571) 272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

TERESA E STRZELECKA
Primary Examiner
Art Unit 1637

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637
September 21, 2015

Notice of References Cited	Application/Control No. 14/540,917	Applicant(s)/Patent Under Reexamination AMORESE ET AL.	
	Examiner TERESA E. STRZELECKA	Art Unit 1637	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2009/0233802 A1	09-2009	Bignell; Helen	C12Q1/6874	506/2
*	B	US-5,384,242 A	01-1995	Oakes; Fred T.	C12Q1/6858	435/6.1
	C	US-				
	D	US-				
	E	US-				
	F	US-				
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Stratagene Catalog, page 39 (1988)
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

IDT Exhibit 1018.0787

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14540917	AMORESE ET AL.
	Examiner	Art Unit
	TERESA E STRZELECKA	1637

CPC- SEARCHED		
Symbol	Date	Examiner
C12Q1/6869 | C12Q2563/179 | C12Q2535/122 |C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 |C12Q2521/319 | C12Q2521/507 | C12Q2525/15 |C12Q2525/155 | C12Q2525/161 | C12Q2525/185 |C12Q2525/191 | C12Q2525/197	5/13/15	/TS/

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
IDS (12/4/14; 3/21/15)	5/13/15	/TS/
WEST	5/13/15	/TS/
Google	5/13/15	/TS/
Inventor name search	5/13/15	/TS/
IDS (8/19/15)	9/21/15	/TS/

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner

IDT Exhibit 1018.0788

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14540917	AMORESE ET AL.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	05/14/2015	09/21/2015							
	1	✓	✓							
	2	✓	✓							
	3	✓	✓							
	4	N	N							
	5	N	N							
	6	N	N							
	7	✓	✓							
	8	N	N							
	9	✓	✓							
	10	✓	✓							
	11	✓	✓							
	12	✓	✓							
	13	✓	✓							
	14	✓	✓							
	15	✓	✓							
	16	✓	✓							
	17	✓	✓							
	18	✓	✓							
	19	✓	✓							
	20	✓	✓							
	21	✓	✓							

IDT Exhibit 1018.0789

Electronic Acknowledgement Receipt

EFS ID:	24457246
Application Number:	14540917
International Application Number:	
Confirmation Number:	7076
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Customer Number:	58249
Filer:	Jeffrey R. Bolland/virginia melton
Filer Authorized By:	Jeffrey R. Bolland
Attorney Docket Number:	NUGN-001/01US
Receipt Date:	28-DEC-2015
Filing Date:	13-NOV-2014
Time Stamp:	15:12:51
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	$2600
RAM confirmation Number	2261
Deposit Account	501283
Authorized User	COOLEY LLP

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 CFR 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 CFR 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 CFR 1.19 (Document supply fees)

Charge any Additional Fees required under 37 CFR 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	NUGN-001_01US_Cert_and_Req_for_Prioritized_Exam.pdf	135696 0dd147da63eaed0cacb097e357149b12ab182145	no	1
Warnings:					
Information:					
2	Request for Continued Examination (RCE)	NUGN-001_01US_RCE_Transmittal.PDF	1350068 17c1beed1fd398e7fc15527e5b89883d989c2fd6	no	3
Warnings:					
Information:					
3	Response After Final Action	NUGN-001_01US_Resp_to_Final.pdf	149091 f35b95e3f3d84f7c33cc351950f09e597fb555e5	no	13
Warnings:					
Information:					
4	Fee Worksheet (SB06)	fee-info.pdf	32536 c48a8781ae3fd724a6bc85cc9e6f824ed4b7958e	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1667391		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Patent Application Fee Transmittal

Application Number:	14540917
Filing Date:	13-Nov-2014
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Filer:	Jeffrey R. Bolland/virginia melton
Attorney Docket Number:	NUGN-001/01US

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Request for Prioritized Examination	2817	1	2000	2000
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
Request for Continued Examination	2801	1	600	600
		Total in USD ($)		**2600**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:	AMORESE, Douglas, et al.	Confirmation No.:	7076
Application No.:	14/540,917	Group Art Unit:	1637
Filed:	November 13, 2014	Examiner:	Teresa E. Strzelecka

FOR: **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

Filed *via* EFS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT/RESPONSE TO OFFICE ACTION

In response to the Final Office Action dated September 28, 2015, this paper is timely filed within the 3-month period for reply ending December 28, 2015. Please amend the above-identified patent application in the following manner:

Amendments to the Claims are reflected in the listing of the claims which begins on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

IDT Exhibit 1018.0794

REMARKS

I. Status

Claims 1-3 and 7-24 are pending; claims 4-6 have been cancelled; claims 21-24 are new; claim 8 has been withdrawn from consideration; and claims 1-3, 7, and 9-21 have been examined on the merits.

Claims 1, 10, 17, 18, and 21 have been amended. Support for the amendment of claims 1 and 21 can be found in the specification in at least the originally filed claim 17 and paragraphs [0029], [0030], and [0065]. Support for the amendment of claim 10 can be found in the specification in at least paragraph [0040]. Support for the amendment of claims 17 and 18 can be found in the specification in at least paragraphs [0030], [0091], and [0104]. Support for new claims 22 and 23 can be found in the specification in at least paragraphs [0030]. Support for new claim 24 can be found in the specification in at least paragraph [0040].

No new matter is introduced by way of amendment. Accordingly, entry thereof is respectfully requested. Upon entry of this amendment, claims 1-3, 7, and 9-22-24 are under examination.

II. Examiner Interview

Examiner Strzelecka is thanked for taking the time for the Applicant-initiated Examiner Interview held on December 22, 2015. Potential claim amendments were generally discussed, as well as potential arguments for addressing the 35 USC 101 rejection over claim 21.

III. 35 U.S.C. § 112(a)

Claim 21 is rejected as allegedly failing to comply with the written description requirement. With regard to the recitation of "A kit comprising a plurality of double-stranded adaptors…", the Examiner alleges that the specification fails to support the limitation of an adaptor being double-stranded. Office Action, p. 5. Applicant respectfully traverses this rejection.

Paragraph [0047] of the specification recites, "In some embodiments, the adaptor can be a double stranded DNA adaptor." In light of the specification explicitly disclosing the adaptor, in one embodiment, as double-stranded, withdrawal of the rejection is respectfully requested.

IV. 35 U.S.C. § 101

Claim 21 is rejected as allegedly directed to a judicial exception. The Examiner asserts that the claim is drawn to a "plurality of oligonucleotides which have corresponding sequences occurring in nature" and "the additional elements recited in claim 21 do not amount to 'significantly more' than the claimed judicial exceptions of nucleic acid sequences of adapters and primers." Office Action, p. 6. Applicant respectfully traverses this rejection.

Paragraph [0067] of the specification recites, "the identifier sequence is not logically present adjacent to naturally occurring sequence[s]," or in other words, that the combination of the identifier site and target site does not exist in nature. Indeed, if it did, the identifier site could not be used to distinguish between duplicate sequence reads. This runs counter to the Examiner's interpretation of the adaptors as being "sequences occurring in nature". The Office released guidance on 35 U.S.C. § 101 in December 2014 pointing to the patent eligible nature of claims comprising nucleic acids, wherein the nucleic acids comprise a heterologous nucleic acid

sequence. The specification clearly describes the identifier site as not being a part of a naturally occurring sequence. Given that the adaptors of claim 21 necessarily comprise the identifier site, said adaptors are clearly non-naturally occurring. Furthermore, the Examiner has not presented any evidence to support a conclusion that the claimed adaptors have a counterpart in nature. Accordingly, withdrawal of the rejection is respectfully requested.

V. 35 U.S.C. § 103

Claims 1-3, 7, and 9-20 are rejected as allegedly being unpatentable over Samuels *et al.* (US20120220494A1, filed 02/16/12 and published 08/30/12) and McCloskey *et al.* (2007. *Biochem. Genet.* 45:761-767). Office Action, p. 7. Applicant respectfully traverses this rejection.

The Examiner states "Samuels et al. teach a method for detecting a duplicate sequencing read from a population of sample sequence reads." Office Action at p. 7. However, the Examiner also states "Samuels et al. teach detection of PCR-introduced errors in sequencing, ***but do not specifically teach detecting duplicate sequencing reads***." Office Action, p. 8 (emphasis added). Samuels *et al.* discloses a method of providing barcode sequences to targets such as a nucleic acid sequences, wherein universal priming sites may be present in barcode libraries [205], [0216]-[0216]. The barcode sequences and universal priming sites may be attached to the nucleic acid sequences and amplified in a PCR reaction [0288]-[0289]. The amplification products may then be sequenced as a whole or in part [0326-0327]. Sequencing errors versus single nucleotide polymorphisms may be detected through the use of barcodes and sequencing of the amplification products [0339] – [0341].

Samuels *et al.* do not teach (1) the active step of "identifying a duplicate sequencing read as the sequencing read comprising the same identifier site and nucleic acid fragment as another

sequencing read in the population of sequencing reads;" (emphasis added); or (2) adaptors comprising distinct indexing sites and identifier sites. Basic disclosure of general barcodes is not enough to rise to the degree of arriving at the metes and bounds of the instantly claimed combination of an indexing site and an identifier site. Furthermore, Samuels *et al.*is simply not drawn to the detection of duplicate sequencing reads, rather Samuels *et al.* is drawn to the detection of PCR-introduced errors in methods of sequencing.

The Examiner cites paragraphs [0339] – [0341] in support of her assertion that Samuels *et al.* teaches detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target site. A review of paragraphs [0339] – [0341] reveals that the methods of Samuels *et al.* actually disclose a method quite distinct from the instant methods. The disclosed method is as follows. Two fragments are noted to exist in a sample, fragments A and B; and each of fragments A and B have two nucleotide positions of interest, X and Y. Fragments A and B have cytosines at position X, whereas fragment A has an adenine at position Y and fragment B has a guanine at position Y. Fragment A is labelled with barcode 1, and fragment B is labelled with barcode 2. After barcode labelling, each of the fragments are amplified in a single cycle of PCR. Fragment B plus corresponding barcode is replicated correctly, however Fragment A plus corresponding barcode features a base change at position X from cytosine to thymine in one of the strands. The fragments are then subjected to one additional round of PCR. Paragraph [0341] recites a method of comparing the fragments plus corresponding barcodes to identify sequences with the same barcode, but different bases within the fragments; which is contrary to the instant methods of identifying duplications. The methods of Samuels *et al.* are simply not drawn to identifying duplications.

McCloskey *et al.* disclose a method of identifying redundant sequences, based solely on two or more sequence reads sharing an identical "barcode" and "batch-stamp". P. 766, para 4-5. McCloskey *et al.* clearly fails to arrive at the instantly claimed method of "identifying a duplicate sequencing read as the sequencing read comprising the same identifier site and nucleic acid fragment…". (Emphasis added). The indexing site of the instant claims can be considered the "batch-stamp" and the identifier site of the instant can be considered "barcode" solely for the purposes of comparing the instant claims with the disclosure of McCloskey *et al.* However, the instant invention relies on the indexing site for batch identification of samples, and the sequence of the identifier site and the sequence of the fragment that is amplified are used in combination to discern duplication. This is very different than the method disclosed in McCloskey *et al.*

The instant methods are superior to McCloskey *et al.*, as McCloskey *et al.* teach the use of barcodes consisting of 7 nucleotides, which allows for 16,384 unique combinations of adaptors. P. 764, para. 2. This means that in just a single sample comprising 250,000 fragments attached to the "encoding oligonucleotide" of McCloskey *et al.*, there could be as many as 15 sequences that share the exact same barcode and batch-stamp. Operating within the teachings of McCloskey *et al.*, these 15 sequences would be falsely identified as duplicates. Furthermore, it becomes costly to generate 16,384 unique adaptors. Theoretically, the instant methods allow for identifier sites consisting of only 1 nucleotide, which allows for the use of just 4 unique adaptors per sample, because identifying two or more sequences as duplicates in the method of the present invention requires only that the combination of the identifier site and the fragment attached to the adaptor be identical. The lower complexity of adaptors comprising fewer nucleotides provides for cost-effective production of adaptors while maintaining the ability to correctly identify true duplications.

There are several other aspects of the claimed invention that clearly distinguish over McCloskey *et al.* For example, the methods of McCloskey *et al.* only work on single stranded polynucleotides when polymerase extension can be performed at non-thermophilic temperatures, because the oligo used as the primer requires secondary structure to keep it from interfering in the reaction. In contrast to the single stranded annealed introduction of McCloskey *et al.*, the present invention employs a double stranded ligated introduction to the reaction. Further, McCloskey describes the use of an oligo that contains a sequence specific portion for targeting of specific nucleic acid sequences. The adapter employed in the claimed invention does not have any target specific regions and is more economical because it can be used in combination with any target.

In another aspect of the claimed invention, new claims 22 and 23 recite "the sequence of the identifier site is not variable in a plurality of adaptors", which reads on a population of the plurality of adaptors sharing the same identifier site. In contrast, the methods of McCloskey *et al.* require each of the identifier sites or "barcodes" to be unique. Furthermore, McCloskey's "batch stamp" is used solely for the purose of detecting contaminants in a given reaction. In contrast, the indexing site of the present invention is used to separate out data from different samples after multiplex amplification (captured in new claim 24). This use of the indexing site in combination with the identifier site further highlights the deficiencies of McCloskey et al., in that it would be extremely cost prohibitive and difficult to design the many thousands of different primers that would be required to ensure that all primers are unique, as the methods of McCloskey et al. would require.

In sum, the teachings of McCloskey *et al.* fail to adequately address the deficiencies of Samuels *et al.* in arriving at a *prima facie* case for obviousness in the instant claims. Samuels *et*

al. is drawn to methods that run counter to the instantly claimed methods as discussed above, and McCloskey *et al.* utilizes a method that is very different than the presently claimed methods. Withdrawal of the rejection is respectfully requested.

Claim 21 is rejected as allegedly being unpatentable over Samuels *et al.* and McCloskey *et al.*, Bignell *et al.* (US2009/0233802A1; published September 2009), Oakes (US Patent 5,384,242A), and Stratagene Catalog (1998. p. 39). Office Action, p. 9. Withdrawal of the rejection is respectfully requested.

The Examiner asserts that the combination of Samuels *et al.* and McCloskey *et al.* arrives at the double-stranded adapters, and further presents the disclosures of Bignell *et al.,* Oakes, and Stratagene to arrive at a kit and the components thereof that are not disclosed by Samuels *et al.* and McCloskey *et al.* As previously presented, the disclosures of Samuels *et al.* and McCloskey *et al.* fail to arrive at the instantly claimed adaptors of the claimed kit. McCloskey *et al.* is necessarily limited to single-stranded adaptors, and Samuels *et al.* fail to disclose methods utilizing adaptors comprising distinct indexing sites and identifier sites as claimed Withdrawal of the rejection is respectfully requested.

CONCLUSION

In view of the foregoing, Applicant respectfully submits that no further impediments exist to the allowance of this application and, therefore, requests an indication of allowability. However, the Examiner is requested to call the undersigned if any questions or comments arise.

The Director is hereby authorized to charge any appropriate fees under 37 C.F.R. §§1.16, 1.17, and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 50-1283.

Dated: December 28, 2015

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (202) 728-7017
Fax: (202) 842-7899

By: /Jeffrey R. Bolland/
Jeffrey R. Bolland, Ph.D.
Reg. No. 73,759

CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 CFR 1.102(e) (Page 1 of 1)

First Named Inventor:	AMORESE, Douglas	Nonprovisional Application Number (if known):	14/540,917
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1) and the prioritized examination fee set forth in 37 CFR 1.17(c) have been filed with the request. The publication fee requirement is met because that fee, set forth in 37 CFR 1.18(d), is currently $0. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.

2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims, and that any request for an extension of time will cause an outstanding Track I request to be dismissed.

3. The applicable box is checked below:

I. ☐ Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)

i. (a) The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.

---OR---

(b) The application is an original nonprovisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.

ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, **or** the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. ☒ Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)

i. A request for continued examination has been filed with, or prior to, this form.
ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
iii. The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature /Jeffrey R. Bolland/	Date December 28, 2015
Name (Print/Typed) Jeffrey R. Bolland	Practitioner Registration Number 73759

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

☒ *Total of 1 forms are submitted.

American LegalNet, Inc.
www.FormsWorkFlow.com

Doc code: RCEX
Doc description: Request for Continued Examination (RCE)

PTO/SB/30EFS (07-14)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

REQUEST FOR CONTINUED EXAMINATION(RCE)TRANSMITTAL
(Submitted Only via EFS-Web)

Application Number	14/540,917	Filing Date	2014-11-13	Docket Number (if applicable)	NUGN-001/01US	Art Unit	1637
First Named Inventor	Douglas AMORESE			Examiner Name	Teresa E. Strzelecka		

This is a Request for Continued Examination (RCE) under 37 CFR 1.114 of the above-identified application.
Request for Continued Examination (RCE) practice under 37 CFR 1.114 does not apply to any utility or plant application filed prior to June 8 1995, to any international application that does not comply with the requirements of 35 U.S.C. 371, or to any design application. The Instruction Sheet for this form is located at WWW.USPTO.GOV.

SUBMISSION REQUIRED UNDER 37 CFR 1.114

Note: If the RCE is proper, any previously filed unentered amendments and amendments enclosed with the RCE will be entered in the orde in which they were filed unless applicant instructs otherwise. If applicant does not wish to have any previously filed unentered amendment(s entered, applicant must request non-entry of such amendment(s).

☐ Previously submitted. If a final Office action is outstanding, any amendments filed after the final Office action may be considered as a submission even if this box is not checked.

- ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
- ☐ Other _____

☒ Enclosed

- ☒ Amendment/Reply
- ☐ Information Disclosure Statement (IDS)
- ☐ Affidavit(s)/ Declaration(s)
- ☐ Other _____

MISCELLANEOUS

☐ Suspension of action on the above-identified application is requested under 37 CFR 1.103(c) for a period of months _____ (Period of suspension shall not exceed 3 months; Fee under 37 CFR 1.17(i) required)

☐ Other _____

FEES

The RCE fee under 37 CFR 1.17(e) is required by 37 CFR 1.114 when the RCE is filed.

☒ The Director is hereby authorized to charge any underpayment of fees, or credit any overpayments, to Deposit Account No 501283

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

☒ Patent Practitioner Signature

Applicant Signature

IDT Exhibit 1018.0804

PTO/SB/30EFS (07-14)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Signature of Registered U.S. Patent Practitioner			
Signature	/Jeffrey R. Bolland/	Date (YYYY-MM-DD)	2015-12-28
Name	Jeffrey R. Bolland	Registration Number	73759

This collection of information is required by 37 CFR 1.114. The information is required to obtain or retain a benefit by the public which is t file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Tim will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions fo reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

IDT Exhibit 1018.0805

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the informatio solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record ma be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IN THE CLAIMS:

1. (Currently Amended) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

    a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

        (i) an indexing primer binding site;

        (ii) an indexing site;

        (iii) an identifier site consisting of~~comprising at least~~ between 1 and 8 nucleotides; and

        (iv) a target sequence primer binding site;

    b) amplifying the adapter-nucleic acid fragment ligated products;

    c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

    d) ~~detecting~~ identifying a duplicate sequencing read as the ~~population of~~ sequencing read~~s~~ comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads~~a sequencing read with a duplicate identifier site and target sequence~~;

    wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is ~~not conserved across all adaptor sequences~~variable in a plurality of adaptors.

2. (Original) The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
3. (Original) The method of claim 1, wherein the identifier site is sequenced with the indexing site.
4. (Cancelled)
5. (Cancelled)
6. (Cancelled)
7. (Original) The method of claim 1, wherein the adaptor comprises from 5' to 3':
    (i) the indexing primer binding site;
    (ii) the indexing site;
    (iii) the identifier site; and
    (iv) the target sequence primer binding site.
8. (Withdrawn) The method of claim 1, wherein the adaptor comprises from 5' to 3':
    (i) the indexing primer binding site;
    (ii) the indexing site;
    (iii) the target sequence primer binding site; and
    (iv) the identifier site.
9. (Original) The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.
10. (Currently Amended) The method of claim 9, wherein the nucleic acid fragments from each sample [[has]]<u>have</u> the same indexing site<u>, and each sample has a unique indexing site</u>.

11. (Original) The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. (Original) The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. (Original) The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. (Original) The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. (Original) The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. (Original) The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. (Currently Amended) The method of claim 1, wherein the identifier site is ~~at least 10~~ 6 nucleotides in length.

18. (Currently Amended) The method of claim 1, wherein the identifier site is ~~about~~ 8 nucleotides in length.

19. (Original) The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. (Original) The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

21. (Currently Amended) A kit comprising a plurality of double-stranded adaptors, wherein each adaptor comprises:

(i) an indexing primer binding site;

(ii) an indexing site;

(iii) an identifier site ~~comprising at least~~ consisting of between 1 and 8 nucleotides; and

(iv) a target sequencing primer binding site;

wherein the kit further comprises one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained; wherein the indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is ~~not conserved across all adaptor sequences~~variable in a plurality of adaptors.

22. (New) The method of claim 1, wherein the sequence of the identifier site is not variable in a plurality of adaptors.

23. (New) The kit of claim 21, wherein the sequence of the identifier site is not variable in a plurality of adaptors.

24. (New) The method of claim 10, wherein the method further comprises separating fragments or sequencing reads based on indexing site.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/540,917	Filing Date 11/13/2014	☐ To be Mailed

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

	(Column 1)	(Column 2)		
FOR	NUMBER FILED	NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

		(Column 1)		(Column 2)	(Column 3)		
AMENDMENT	12/28/2015	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 21	Minus	** 21	= 0	X $80 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	***3	= 0	X $420 =	0
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	**0**

		(Column 1)		(Column 2)	(Column 3)		
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
/CORALIA BETANCOURT/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

IDT Exhibit 1018.0811

Cooley

FACSIMILE

TO:	Phone Number:	Fax Number:
USPTO - Examiner Strzelecka		**571-273-0789**

FROM:		*FOR COOLEY USE ONLY*
Jeffrey R. Bolland		**REQUESTOR NUMBER:**
Sender's Direct Line:	+1 202 728 7017	21175
Reply Fax Number:	+1 202 842-7899	**CLIENT MATTER:**
Total Pages Including Cover:	[5]	301735-2003

DATE: December 22, 2015

RE: **U.S. Application Serial No. 14/540,917**
Our Ref. No.: NUGN-001/01US (301735-2003)

COMMENTS:

Please see the attached claims in connection with U.S. Application Serial No. 14/540,917.

If you do not receive all of the pages or find that they are illegible, please call Sender's Name at Sender's Phone Number.

COOLEY LLP 1299 PENNSYLVANIA AVENUE, NW, SUITE 700 WASHINGTON, DC 20004-2400
T: (202) 842-7800 F: (202) 842-7899 COOLEY.COM

PAGE 1/5 * RCVD AT 12/22/2015 1:10:32 PM [Eastern Standard Time] * SVR:W-PTOFAX-001/45 * DNIS:2730789 * CSID: * DURATION (mm-ss):01-32

FOR DISCUSSION PURPOSES ONLY

IN THE CLAIMS:

1. (Currently Amended) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

   a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

      (i) an indexing primer binding site;

      (ii) an indexing site;

      (iii) an identifier site consisting of~~comprising at least~~ between 1 and 8 nucleotides; and

      (iv) a target sequence primer binding site;

   b) amplifying the adapter-nucleic acid fragment ligated products;

   c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

   d) ~~detecting~~ identifying a duplicate sequencing read as the ~~population of~~ sequencing reads comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads~~a sequencing read with a duplicate identifier site and target sequence~~;

wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is ~~not conserved across all adaptor sequences~~variable in a plurality of adaptors.

Cooley LLP DRAFT DATED: DECEMBER 17, 2015

PAGE 2/5 * RCVD AT 12/22/2015 1:10:32 PM [Eastern Standard Time] * SVR:W-PTOFAX-001/45 * DNIS:2730789 * CSID: * DURATION (mm-ss):01-32

IDT Exhibit 1018.0813

2. (Original) The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
3. (Original) The method of claim 1, wherein the identifier site is sequenced with the indexing site.
4. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the indexing site.
5. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced with the target sequence.
6. (Withdrawn) The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.
7. (Original) The method of claim 1, wherein the adaptor comprises from 5' to 3':
    (i) the indexing primer binding site;
    (ii) the indexing site;
    (iii) the identifier site; and
    (iv) the target sequence primer binding site.
8. (Withdrawn) The method of claim 1, wherein the adaptor comprises from 5' to 3':
    (i) the indexing primer binding site;
    (ii) the indexing site;
    (iii) the target sequence primer binding site; and
    (iv) the identifier site.
9. (Original) The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.

Cooley LLP DRAFT DATED: DECEMBER 17, 2015

PAGE 3/5 * RCVD AT 12/22/2015 1:10:32 PM [Eastern Standard Time] * SVR:W-PTOFAX-001/45 * DNIS:2730789 * CSID: * DURATION (mm-ss):01-32

10. (Currently Amended) The method of claim 9, wherein the nucleic acid fragments from each sample [[has]]have the same indexing site, and each sample has a unique indexing site.

11. (Original) The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. (Original) The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. (Original) The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. (Original) The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. (Original) The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. (Original) The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. (Currently Amended) The method of claim 1, wherein the identifier site is ~~at least 10~~ 6 nucleotides in length.

18. (Currently Amended) The method of claim 1, wherein the identifier site is ~~about~~ 8 nucleotides in length.

19. (Original) The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. (Original) The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

Cooley LLP DRAFT DATED: DECEMBER 17, 2015

PAGE 4/5 * RCVD AT 12/22/2015 1:10:32 PM [Eastern Standard Time] * SVR:W-PTOFAX-001/45 * DNIS:2730789 * CSID: * DURATION (mm-ss):01-32

IDT Exhibit 1018.0815

21. (Currently Amended) A kit comprising a plurality of double-stranded adaptors, wherein each adaptor comprises:

(i) an indexing primer binding site;

(ii) an indexing site;

(iii) an identifier site ~~comprising at least~~ consisting of between 1 and 8 nucleotides; and

(iv) a target sequencing primer binding site;

wherein the kit further comprises one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained; wherein the indexing site is unique amongst a subset of the adaptors and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is ~~not conserved across all adaptor sequences~~variable in a plurality of adaptors.

22. (New) The method of claim 1, wherein the sequence of the identifier site is not variable in a plurality of adaptors.

23. (New) The kit of claim 21, wherein the sequence of the identifier site is not variable in a plurality of adaptors.

24. (New) The method of claim 10, wherein the method further comprises separating fragments or sequencing reads based on indexing site.

Cooley LLP
DRAFT
DATED: DECEMBER 17, 2015

PAGE 5/5 * RCVD AT 12/22/2015 1:10:32 PM [Eastern Standard Time] * SVR:W-PTOFAX-001/45 * DNIS:2730789 * CSID: * DURATION (mm-ss):01-32

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/540,917	11/13/2014	Douglas AMORESE	NUGN-001/01US	7076

58249 7590 01/14/2016

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue, NW
Suite 700
Washington, DC 20004

EXAMINER
STRZELECKA, TERESA E

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
01/14/2016	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zpatdcdocketing@cooley.com

IDT Exhibit 1018.0817

Applicant-Initiated Interview Summary	Application No.	Applicant(s)	
	14/540,917	AMORESE ET AL.	
	Examiner	Art Unit	
	TERESA E. STRZELECKA	1637	

All participants (applicant, applicant's representative, PTO personnel):

(1) *Teresa Strzelecka*.

(2) *Jeffrey Bolland*.

(3) *Bonnie McLeod*.

(4) _____.

Date of Interview: *22 December 2015*.

Type: ☒ Telephonic ☐ Video Conference
☐ Personal [copy given to: ☐ applicant ☐ applicant's representative]

Exhibit shown or demonstration conducted: ☐ Yes ☒ No.
If Yes, brief description: _____.

Issues Discussed ☐101 ☐112 ☐102 ☐103 ☒Others
(For each of the checked box(es) above, please describe below the issue and detailed description of the discussion)

Claim(s) discussed: *1*.

Identification of prior art discussed: _____.

Substance of Interview
(For each issue discussed, provide a detailed description and indicate if agreement was reached. Some topics may include: identification or clarification of a reference or a portion thereof, claim interpretation, proposed amendments, arguments of any applied references etc...)

Applicant's representaive submitted proposed claim amendments which were discussed during the interview.

Applicant recordation instructions: The formal written reply to the last Office action must include the substance of the interview. (See MPEP section 713.04). If a reply to the last Office action has already been filed, applicant is given a non-extendable period of the longer of one month or thirty days from this interview date, or the mailing date of this interview summary form, whichever is later, to file a statement of the substance of the interview

Examiner recordation instructions: Examiners must summarize the substance of any interview of record. A complete and proper recordation of the substance of an interview should include the items listed in MPEP 713.04 for complete and proper recordation including the identification of the general thrust of each argument or issue discussed, a general indication of any other pertinent matters discussed regarding patentability and the general results or outcome of the interview, to include an indication as to whether or not agreement was reached on the issues raised.

☒ Attachment

/TERESA E STRZELECKA/ Primary Examiner, Art Unit 1637	

IDT Exhibit 1018.0818

Summary of Record of Interview Requirements

Manual of Patent Examining Procedure (MPEP), Section 713.04, Substance of Interview Must be Made of Record
A complete written statement as to the substance of any face-to-face, video conference, or telephone interview with regard to an application must be made of record in the application whether or not an agreement with the examiner was reached at the interview.

Title 37 Code of Federal Regulations (CFR) § 1.133 Interviews
Paragraph (b)

In every instance where reconsideration is requested in view of an interview with an examiner, a complete written statement of the reasons presented at the interview as warranting favorable action must be filed by the applicant. An interview does not remove the necessity for reply to Office action as specified in §§ 1.111, 1.135. (35 U.S.C. 132)

37 CFR §1.2 Business to be transacted in writing.
All business with the Patent or Trademark Office should be transacted in writing. The personal attendance of applicants or their attorneys or agents at the Patent and Trademark Office is unnecessary. The action of the Patent and Trademark Office will be based exclusively on the written record in the Office. No attention will be paid to any alleged oral promise, stipulation, or understanding in relation to which there is disagreement or doubt.

———

The action of the Patent and Trademark Office cannot be based exclusively on the written record in the Office if that record is itself incomplete through the failure to record the substance of interviews.

It is the responsibility of the applicant or the attorney or agent to make the substance of an interview of record in the application file, unless the examiner indicates he or she will do so. It is the examiner's responsibility to see that such a record is made and to correct material inaccuracies which bear directly on the question of patentability.

Examiners must complete an Interview Summary Form for each interview held where a matter of substance has been discussed during the interview by checking the appropriate boxes and filling in the blanks. Discussions regarding only procedural matters, directed solely to restriction requirements for which interview recordation is otherwise provided for in Section 812.01 of the Manual of Patent Examining Procedure, or pointing out typographical errors or unreadable script in Office actions or the like, are excluded from the interview recordation procedures below. Where the substance of an interview is completely recorded in an Examiners Amendment, no separate Interview Summary Record is required.

The Interview Summary Form shall be given an appropriate Paper No., placed in the right hand portion of the file, and listed on the "Contents" section of the file wrapper. In a personal interview, a duplicate of the Form is given to the applicant (or attorney or agent) at the conclusion of the interview. In the case of a telephone or video-conference interview, the copy is mailed to the applicant's correspondence address either with or prior to the next official communication. If additional correspondence from the examiner is not likely before an allowance or if other circumstances dictate, the Form should be mailed promptly after the interview rather than with the next official communication.

The Form provides for recordation of the following information:
- Application Number (Series Code and Serial Number)
- Name of applicant
- Name of examiner
- Date of interview
- Type of interview (telephonic, video-conference, or personal)
- Name of participant(s) (applicant, attorney or agent, examiner, other PTO personnel, etc.)
- An indication whether or not an exhibit was shown or a demonstration conducted
- An identification of the specific prior art discussed
- An indication whether an agreement was reached and if so, a description of the general nature of the agreement (may be by attachment of a copy of amendments or claims agreed as being allowable). Note: Agreement as to allowability is tentative and does not restrict further action by the examiner to the contrary.
- The signature of the examiner who conducted the interview (if Form is not an attachment to a signed Office action)

It is desirable that the examiner orally remind the applicant of his or her obligation to record the substance of the interview of each case. It should be noted, however, that the Interview Summary Form will not normally be considered a complete and proper recordation of the interview unless it includes, or is supplemented by the applicant or the examiner to include, all of the applicable items required below concerning the substance of the interview.

A complete and proper recordation of the substance of any interview should include at least the following applicable items:
1) A brief description of the nature of any exhibit shown or any demonstration conducted,
2) an identification of the claims discussed,
3) an identification of the specific prior art discussed,
4) an identification of the principal proposed amendments of a substantive nature discussed, unless these are already described on the Interview Summary Form completed by the Examiner,
5) a brief identification of the general thrust of the principal arguments presented to the examiner,
   (The identification of arguments need not be lengthy or elaborate. A verbatim or highly detailed description of the arguments is not required. The identification of the arguments is sufficient if the general nature or thrust of the principal arguments made to the examiner can be understood in the context of the application file. Of course, the applicant may desire to emphasize and fully describe those arguments which he or she feels were or might be persuasive to the examiner.)
6) a general indication of any other pertinent matters discussed, and
7) if appropriate, the general results or outcome of the interview unless already described in the Interview Summary Form completed by the examiner.

Examiners are expected to carefully review the applicant's record of the substance of an interview. If the record is not complete and accurate, the examiner will give the applicant an extendable one month time period to correct the record.

Examiner to Check for Accuracy

If the claims are allowable for other reasons of record, the examiner should send a letter setting forth the examiner's version of the statement attributed to him or her. If the record is complete and accurate, the examiner should place the indication, "Interview Record OK" on the paper recording the substance of the interview along with the date and the examiner's initials.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

MAILED
FEB 12 2016
OFFICE OF PETITIONS

COOLEY LLP
ATTN: PATENT GROUP
1299 PENNSYLVANIA AVNEUE, NW
SUITE 700
WASHINGTON DC 20004

Doc Code: TRACK1.GRANT

Decision Granting Request for Prioritized Examination (Track I or After RCE)	Application No.: 14/540,917

1. THE REQUEST FILED 12/28/15 IS **GRANTED.**

The above-identified application has met the requirements for prioritized examination

A. ☐ for an original nonprovisional application (Track I).
B. ☒ for an application undergoing continued examination (RCE).

2. **The above-identified application will undergo prioritized examination.** The application will be accorded special status throughout its entire course of prosecution until one of the following occurs:

A. filing a **petition for extension of time** to extend the time period for filing a reply;
B. filing an **amendment to amend the application to contain more than four independent claims, more than thirty total claims**, or a multiple dependent claim;
C. filing a **request for continued examination**;
D. filing a notice of appeal;
E. filing a request for suspension of action;
F. mailing of a notice of allowance;
G. mailing of a final Office action;
H. completion of examination as defined in 37 CFR 41.102; or
I. abandonment of the application.

Telephone inquiries with regard to this decision should be directed to Cheryl Gibson-Baylor at (571)272-3213, Office of Petitions. In his/her absence, calls may be directed to Brian W. Brown, (571)272-5338.

Cheryl Gibson-Baylor
/Cheryl Gibson-Baylor/
[*Signature*]

Petitions Paralegal Specialist
(Title)

IDT Exhibit 1018.0820

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/540,917	11/13/2014	Douglas AMORESE	NUGN-001/01US	7076

58249 7590 03/03/2016

COOLEY LLP
ATTN: Patent Group
1299 Pennsylvania Avenue, NW
Suite 700
Washington, DC 20004

EXAMINER	
STRZELECKA, TERESA E	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
03/03/2016	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zpatdcdocketing@cooley.com

IDT Exhibit 1018.0821

Office Action Summary	Application No. 14/540,917	Applicant(s) AMORESE ET AL.	
	Examiner TERESA E. STRZELECKA	Art Unit 1637	AIA (First Inventor to File) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12/28/15*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-3 and 7-24* is/are pending in the application.

5a) Of the above claim(s) *8* is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-3,7 and 9-24* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

IDT Exhibit 1018.0822

1. The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

2. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on December 28, 2015 has been entered.

3. Claims 1-21 were previously pending, with claims 4-6 and 8 withdrawn from consideration. Applicant cancelled claims 4-6, amended claims 1, 10, 17, 18 and 21 and added new claims 22-24. Claims 1-3, 7 and 9-24 are under consideration.

4. Applicant's amendments overcame the previously presented rejections with the exception of the rejection of claim 21 under 35 U.S.C. 101. Arguments regarding this rejection are addressed in the "Response to Arguments" section below.

5. This office action contains new grounds for rejection necessitated by amendment.

Response to Arguments

6. Applicant's arguments filed December 28, 2015 have been fully considered but they are not persuasive.

Regarding the rejection of claim 21 under 35 U.S.C. 101, Applicant argues the following:

"Paragraph [0067] of the specification recites, "the identifier sequence is not logically present adjacent to naturally occurring sequence[s]," or in other words, that the combination of

the identifier site and target site does not exist in nature. Indeed, if it did, the identifier site could not be used to distinguish between duplicate sequence reads. This runs counter to the Examiner's interpretation of the adaptors as being "sequences occurring in nature". The Office released guidance on 35 U.S.C. § 101 in December 2014 pointing to the patent eligible nature of claims comprising nucleic acids, wherein the nucleic acids comprise a heterologous nucleic acid sequence. The specification clearly describes the identifier site as not being a part of a naturally occurring sequence. Given that the adaptors of claim 21 necessarily comprise the identifier site, said adaptors are clearly non-naturally occurring. Furthermore, the Examiner has not presented any evidence to support a conclusion that the claimed adaptors have a counterpart in nature. Accordingly, withdrawal of the rejection is respectfully requested."

First, even assuming that one would understand what "...not logically present adjacent to naturally occurring sequence…" means, almost any sequence is bound to be found in nature. The claim does not limit the length of the adaptors, therefore, for example, they can be 8 bp long. As detailed by Liu et al. (BMC Genomics, vol. 9:509, pp. 1-17, 2008), almost all possible 12mers are present in a human genome, and other genomes contain variable numbers of the different n-mers (Abstract; page 2, paragraphs 4-6; page 3, page 5). Further, considering the fact that in 2013 GenBank database contained 4.5×10^{11} bp of sequences (Benson et al., Nucl. Acids Res., vol. 41, D36-D42, 2013; see Table 1), and that there are 4^8, or 6.5×10^4 possible 8mer sequences, each one of them would be represented multiple times in the GenBank database, as would every possible 16mer, for example, of which there are 4.3×10^9 possible sequences.

Finally, Applicant's sequence with SEQ ID NO: 4, which has 58 bp without the 3' 8-bp random part was found in the GenBank database 106 times (see attached BLAST search result).

In conclusion, the claim is not patent eligible under 35 U.S.C. 101. The rejection is maintained.

Claim Rejections - 35 USC § 112

7. The following is a quotation of 35 U.S.C. 112(b):

> (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 1-3, 7 and 9-24 rejected under 35 U.S.C. 112(b) or 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor, or for pre-AIA the applicant regards as the invention.

A) Claims 1, 7, 9-20, 22 and 24 are indefinite in claim 1. Claim 1 is indefinite over the recitation of "…wherein the sequence of the identifier site is variable in a plurality of adaptors". It is not clear what the term "variable" means: does it mean sequence length variability or sequence content variability? In conclusion, the claims do not have clear metes and bounds.

B) Claim 22 is indefinite over the recitation of "…wherein the sequence of the identifier site is not variable in a plurality of adaptors". First, it is not clear what the term "variable" means: does it mean sequence length variability or sequence content variability? Further, claim

22 depends from claim 1, which requires that "…the sequence of the identifier site is variable in a plurality of adaptors." It is therefore not clear how a sequence identifier can be both variable and not variable in the same plurality of adaptors.

In conclusion, the claim does not have clear metes and bounds.

C) Claims 21 and 23 are indefinite in claim 21. Claim 21 is indefinite over the recitation of "…wherein the sequence of the identifier site is variable in a plurality of adaptors". It is not clear what the term "variable" means: does it mean sequence length variability or sequence content variability? In conclusion, the claims do not have clear metes and bounds.

D) Claim 23 is indefinite over the recitation of "…wherein the sequence of the identifier site is not variable in a plurality of adaptors". First, it is not clear what the term "variable" means: does it mean sequence length variability or sequence content variability? Further, claim 23 depends from claim 21, which requires that "…the sequence of the identifier site is variable in a plurality of adaptors." It is therefore not clear how a sequence identifier can be both variable and not variable in the same plurality of adaptors.

In conclusion, the claim does not have clear metes and bounds.

Claim Rejections - 35 USC § 101

9. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

10. Claims 21 and 23 are rejected under 35 U.S.C. 101 because the claimed invention is directed to a judicial exception (i.e., a law of nature, a natural phenomenon, or an abstract idea) without significantly more.

Claim(s) 21 is directed to a kit comprising a plurality of adaptors, i.e., nucleic acid sequences, comprising an indexing primer binding site, an indexing site, an identifier site and a target sequencing binding site, the kit further comprising one or more oligonucleotide primers, reagents for amplification, and instructions for employing the kit; wherein the plurality of adaptors, one or more oligonucleotide primers, and reagents for amplification are separately contained.

No specific sequences of the adapters and oligonucleotide primers are claimed; therefore the claim is drawn to a plurality of oligonucleotides which have corresponding sequences occurring in nature. The claim does not limit the length of the adaptors, therefore, for example, they can be 8 bp long. As detailed by Liu et al. (BMC Genomics, vol. 9:509, pp. 1-17, 2008), almost all possible 12mers are present in a human genome, and other genomes contain variable numbers of the different n-mers (Abstract; page 2, paragraphs 4-6; page 3, page 5). Further, considering the fact that in 2013 GenBank database contained 4.5×10^{11} bp of sequences (Benson et al., Nucl. Acids Res., vol. 41, D36-D42, 2013; see Table 1), and that there are 4^8, or 6.5×10^4 possible 8mer sequences, each one of them would be represented multiple times in the GenBank database, as would every possible 16mer, for example, of which there are 4.3×10^9 possible sequences.

Finally, Applicant's sequence with SEQ ID NO: 4, which has 58 bp without the 3' 8bp random part was found in the GenBank database 106 times (see attached BLAST search result).

The claim(s) does/do not include additional elements that are sufficient to amount to significantly more than the judicial exception because the other elements of the kit are well-understood, routine and conventional in the art. Specifically, "reagents for amplification" are not specified, therefore could encompass water and salt solutions, for example. It is conventional in the art to include instructions for using a kit, as well as to partition components of a kit into separate containers (see, for example, US 2009/0233802 A1, [0103]-[0107]; US 5,384,242 A; issued January 1995; col. 11, lines 61-68; col. 12, lines 1-13). Therefore, the additional elements recited in claim 21 do not amount to "significantly more" than the claimed judicial exceptions of nucleic acid sequences of adapters and primers.

In conclusion, claim 21 is not patent eligible under 35 U.S.C. 101.

Claim Rejections - 35 USC § 102

10. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a)(1) the claimed invention was patented, described in a printed publication, or in public use, on sale or otherwise available to the public before the effective filing date of the claimed invention.

11. Claims 21 and 23 are rejected under 35 U.S.C. 102(a)(1) as being anticipated by Casbon et al. (US 2012/0071331 A1).

Regarding claims 21 and 23, Casbon et al. teach a kit comprising in separate containers reaction reagents comprising adapters containing indexing sites, variable identifier sites and primer sites, primers, polymerases, ligases, etc. (page 3, [0021]; page 13, [0091]-[0093]; page 1,

[0005]-[0006]; page 2, [0016]-[0017]; page 3, [0023]; page 7, [0045]-[0046]; page 8, [0051]-[0055]; page 9, [0057]-[0061]).

12. No claims are allowed.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to TERESA E. STRZELECKA whose telephone number is (571)272-0789. The examiner can normally be reached on M-F (8:30-5:30).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on (571) 272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

TERESA E STRZELECKA
Primary Examiner
Art Unit 1637

/TERESA E STRZELECKA/
Primary Examiner, Art Unit 1637
February 25, 2016

Notice of References Cited	Application/Control No. 14/540,917	Applicant(s)/Patent Under Reexamination AMORESE ET AL.	
	Examiner TERESA E. STRZELECKA	Art Unit 1637	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2012/0071331 A1	03-2012	Casbon; James	C12Q1/6855	506/4
	B	US-				
	C	US-				
	D	US-				
	E	US-				
	F	US-				
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Liu, Z. et al., BMC Genomics, vol. 9:509, pp. 1-17 (2008)
	V	Benson, D.A. et al., Nucl. Acids Res., vol. 41, D36-D42 (2013)
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

IDT Exhibit 1018.0831

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14540917	AMORESE ET AL.
	Examiner	**Art Unit**
	TERESA E STRZELECKA	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	05/14/2015	09/21/2015	02/25/2016						
	1	✓	✓	✓						
	2	✓	✓	✓						
	3	✓	✓	✓						
	4	N	N	-						
	5	N	N	-						
	6	N	N	-						
	7	✓	✓	✓						
	8	N	N	N						
	9	✓	✓	✓						
	10	✓	✓	✓						
	11	✓	✓	✓						
	12	✓	✓	✓						
	13	✓	✓	✓						
	14	✓	✓	✓						
	15	✓	✓	✓						
	16	✓	✓	✓						
	17	✓	✓	✓						
	18	✓	✓	✓						
	19	✓	✓	✓						
	20	✓	✓	✓						
	21	✓	✓	✓						
	22			✓						
	23			✓						
	24			✓						

IDT Exhibit 1018.0832

BLAST ®

Basic Local Alignment Search Tool

NCBI/ BLAST/ blastn suite/ **Formatting Results - CTN8H65N014**
Formatting options
Download
Blast report description

SEQID4

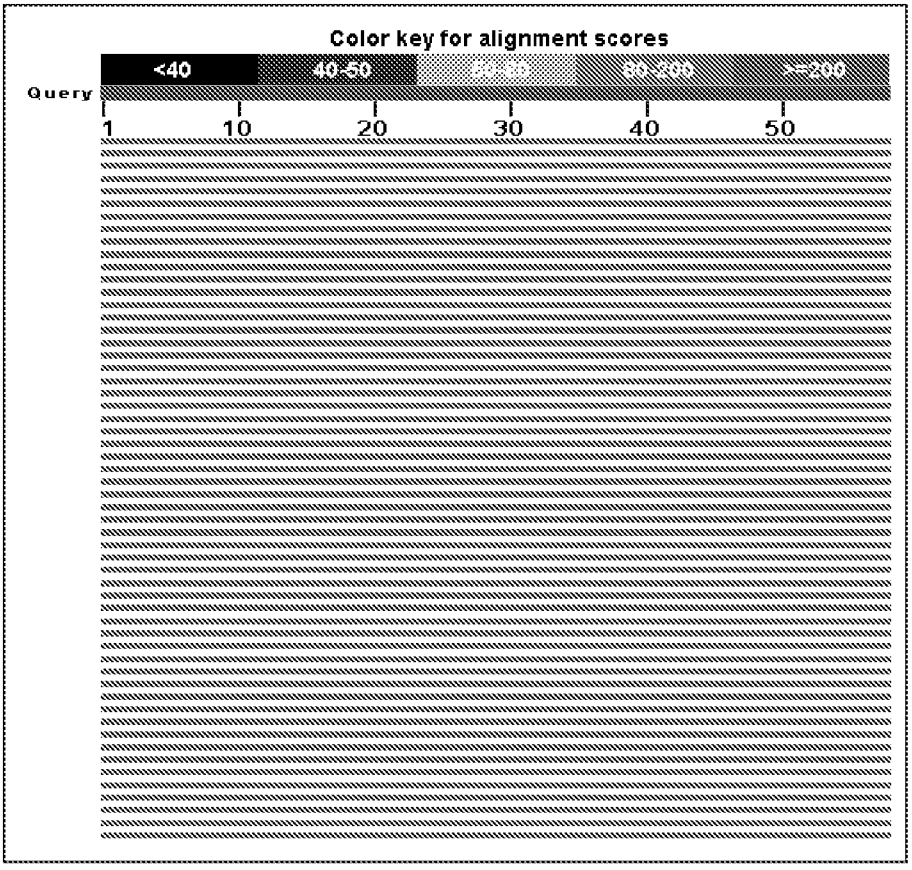
RID CTN8H65N014 (Expires on 02-25 04:54 am)

Query ID lcl|Query_78557
Description None
Molecule type nucleic acid
Query Length 58

Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.3.1+

Graphic Summary

Distribution of 106 Blast Hits on the Query Sequence



Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Clupea harengus uncharacterized LOC105905600	105	105	100%	4e-20	100%	XR_001162831.1

IDT Exhibit 1018.0833

Description	Max score	Total score	Query cover	E value	Ident	Accession
(LOC105905600), transcript variant X11, ncRNA						
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X10, ncRNA	105	105	100%	4e-20	100%	XR_001162630.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X9, ncRNA	105	105	100%	4e-20	100%	XR_001162629.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X8, ncRNA	105	105	100%	4e-20	100%	XR_001162628.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X7, ncRNA	105	105	100%	4e-20	100%	XR_001162627.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X6, ncRNA	105	105	100%	4e-20	100%	XR_001162626.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X5, ncRNA	105	105	100%	4e-20	100%	XR_001162625.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X4, ncRNA	105	105	100%	4e-20	100%	XR_001162624.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X3, ncRNA	105	105	100%	4e-20	100%	XR_001162623.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X2, ncRNA	105	105	100%	4e-20	100%	XR_001162622.1
PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X1, ncRNA	105	105	100%	4e-20	100%	XR_001162621.1
PREDICTED: Clupea harengus uncharacterized LOC105901894 (LOC105901894), transcript variant X2, ncRNA	105	105	100%	4e-20	100%	XR_001162457.1
PREDICTED: Clupea harengus uncharacterized LOC105901894 (LOC105901894), transcript variant X1, ncRNA	105	105	100%	4e-20	100%	XR_001162456.1
PREDICTED: Clupea harengus uncharacterized LOC105898768 (LOC105898768), ncRNA	105	105	100%	4e-20	100%	XR_001162326.1
PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X4, ncRNA	105	105	100%	4e-20	100%	XR_001162263.1
PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X3, ncRNA	105	105	100%	4e-20	100%	XR_001162262.1

IDT Exhibit 1018.0834

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X2, ncRNA	105	105	100%	4e-20	100%	XR_001162281.1
PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X1, ncRNA	105	105	100%	4e-20	100%	XR_001162280.1
PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X3, ncRNA	105	105	100%	4e-20	100%	XR_001162196.1
PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X2, ncRNA	105	105	100%	4e-20	100%	XR_001162195.1
PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X1, ncRNA	105	105	100%	4e-20	100%	XR_001162194.1
Dicrocoelium dendriticum genome assembly D_dendriticum_Leon_v1_0_4, scaffold DDEL_contig0063667	105	105	100%	4e-20	100%	LK629636.1
Nippostrongylus brasiliensis genome assembly N_brasiliensis_RM07_v1_5_4, scaffold NBR_scaffold0003529	105	105	100%	4e-20	100%	LM437148.1
Nippostrongylus brasiliensis genome assembly N_brasiliensis_RM07_v1_5_4, scaffold NBR_scaffold0003502	105	105	100%	4e-20	100%	LM437118.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002276	105	105	100%	4e-20	100%	LN680047.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000045	105	105	100%	4e-20	100%	LN688550.1
Cyprinus carpio genome assembly common carp genome, scaffold 000003581	105	105	100%	4e-20	100%	LN688496.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001750	105	105	100%	4e-20	100%	LN686313.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001003	105	105	100%	4e-20	100%	LN686283.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001567	105	105	100%	4e-20	100%	LN686269.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000916	105	105	100%	4e-20	100%	LN686136.1
Cyprinus carpio genome assembly common carp genome, scaffold 000003735	105	105	100%	4e-20	100%	LN687769.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001344	105	105	100%	4e-20	100%	LN686822.1
	105	186	100%	4e-20	100%	LN686727.1

IDT Exhibit 1018.0835

Description	Max score	Total score	Query cover	E value	Ident	Accession
Cyprinus carpio genome assembly common carp genome, scaffold 000000040						
Cyprinus carpio genome assembly common carp genome, scaffold 000001121	105	105	100%	4e-20	100%	LN596723.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000158	105	105	100%	4e-20	100%	LN596719.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002540	105	105	100%	4e-20	100%	LN596606.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002987	105	105	100%	4e-20	100%	LN596327.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001762	105	105	100%	4e-20	100%	LN596274.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001964	105	211	100%	4e-20	100%	LN596222.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000564	105	105	100%	4e-20	100%	LN596142.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000654	105	105	100%	4e-20	100%	LN595796.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002386	105	105	100%	4e-20	100%	LN595666.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002381	105	105	100%	4e-20	100%	LN595475.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000064	105	105	100%	4e-20	100%	LN595230.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001650	105	105	100%	4e-20	100%	LN595223.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000702	105	105	100%	4e-20	100%	LN595171.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002009	105	105	100%	4e-20	100%	LN595112.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001123	105	105	100%	4e-20	100%	LN595103.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002070	105	105	100%	4e-20	100%	LN594987.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000549	105	105	100%	4e-20	100%	LN594921.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002464	105	105	100%	4e-20	100%	LN594659.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000619	105	105	100%	4e-20	100%	LN594833.1

IDT Exhibit 1018.0836

Description	Max score	Total score	Query cover	E value	Ident	Accession
Cyprinus carpio genome assembly common carp genome, scaffold 000000122	105	314	100%	4e-20	100%	LN594771.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001987	105	105	100%	4e-20	100%	LN594742.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002509	105	105	100%	4e-20	100%	LN594739.1
Cyprinus carpio genome assembly common carp genome, scaffold 000003052	105	105	100%	4e-20	100%	LN594681.1
Cyprinus carpio genome assembly common carp genome, scaffold 000004095	105	105	100%	4e-20	100%	LN594602.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000925	105	105	100%	4e-20	100%	LN594483.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001670	105	105	100%	4e-20	100%	LN594456.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002477	105	105	100%	4e-20	100%	LN594374.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002610	105	105	100%	4e-20	100%	LN594146.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001475	105	105	100%	4e-20	100%	LN594027.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001023	105	105	100%	4e-20	100%	LN593855.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001156	105	105	100%	4e-20	100%	LN593842.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000879	105	105	100%	4e-20	100%	LN593720.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000579	105	105	100%	4e-20	100%	LN593390.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001592	105	105	100%	4e-20	100%	LN593384.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001168	105	105	100%	4e-20	100%	LN593298.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001951	105	105	100%	4e-20	100%	LN592971.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000038	105	105	100%	4e-20	100%	LN592912.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000741	105	105	100%	4e-20	100%	LN592861.1
	105	105	100%	4e-20	100%	LN592771.1

IDT Exhibit 1018.0837

Description	Max score	Total score	Query cover	E value	Ident	Accession
Cyprinus carpio genome assembly common carp genome, scaffold 000002239						
Cyprinus carpio genome assembly common carp genome, scaffold 000001653	105	105	100%	4e-20	100%	LN592703.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001717	105	105	100%	4e-20	100%	LN592632.1
Cyprinus carpio genome assembly common carp genome, scaffold 000006258	105	105	100%	4e-20	100%	LN592578.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002277	105	105	100%	4e-20	100%	LN592402.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001359	105	105	100%	4e-20	100%	LN592387.1
Cyprinus carpio genome assembly common carp genome, scaffold 000008356	105	105	100%	4e-20	100%	LN592384.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000021	105	211	100%	4e-20	100%	LN592208.1
Cyprinus carpio genome assembly common carp genome, scaffold 000000540	105	105	100%	4e-20	100%	LN592009.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001009	105	105	100%	4e-20	100%	LN591795.1
Cyprinus carpio genome assembly common carp genome, scaffold 000002927	105	105	100%	4e-20	100%	LN591675.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001264	105	105	100%	4e-20	100%	LN591563.1
Cyprinus carpio genome assembly common carp genome, scaffold 000007887	105	105	100%	4e-20	100%	LN591511.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001448	105	105	100%	4e-20	100%	LN591505.1
Cyprinus carpio genome assembly common carp genome, scaffold 000029065	105	105	100%	4e-20	100%	LN591261.1
Cyprinus carpio genome assembly common carp genome, scaffold 000029034	105	105	100%	4e-20	100%	LN591245.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028985	105	211	100%	4e-20	100%	LN591218.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028978	105	105	100%	4e-20	100%	LN591212.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028950	105	105	100%	4e-20	100%	LN591199.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028897	105	105	100%	4e-20	100%	LN591168.1

IDT Exhibit 1018.0838

Description	Max score	Total score	Query cover	E value	Ident	Accession
Cyprinus carpio genome assembly common carp genome, scaffold 000028849	105	105	100%	4e-20	100%	LN891137.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028846	105	105	100%	4e-20	100%	LN891136.1
Cyprinus carpio genome assembly common carp genome, scaffold 000028815	105	105	100%	4e-20	100%	LN891118.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001451	105	105	100%	4e-20	100%	LN891088.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001346	105	105	100%	4e-20	100%	LN891081.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001187	105	105	100%	4e-20	100%	LN891036.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001040	105	105	100%	4e-20	100%	LN890987.1
Cyprinus carpio genome assembly common carp genome, scaffold 000001016	105	105	100%	4e-20	100%	LN890978.1

Alignments

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X11, ncRNA
Sequence ID: **ref|XR_001162631.1|** Length: 124 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X10, ncRNA
Sequence ID: **ref|XR_001162630.1|** Length: 133 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X9, ncRNA
Sequence ID: **ref|XR_001162629.1|** Length: 131 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame

IDT Exhibit 1018.0839

105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X8, ncRNA

Sequence ID: **ref|XR_001162628.1|** Length: 141 Number of Matches: 1

Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X7, ncRNA

Sequence ID: **ref|XR_001162627.1|** Length: 137 Number of Matches: 1

Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X6, ncRNA

Sequence ID: **ref|XR_001162626.1|** Length: 140 Number of Matches: 1

Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X5, ncRNA

Sequence ID: **ref|XR_001162625.1|** Length: 139 Number of Matches: 1

Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X4, ncRNA

Sequence ID: **ref|XR_001162624.1|** Length: 133 Number of Matches: 1

Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame

IDT Exhibit 1018.0840

105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X3, ncRNA

Sequence ID: **ref|XR_001162623.1|** Length: 144 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X2, ncRNA

Sequence ID: **ref|XR_001162622.1|** Length: 124 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105905600 (LOC105905600), transcript variant X1, ncRNA

Sequence ID: **ref|XR_001162621.1|** Length: 129 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

PREDICTED: Clupea harengus uncharacterized LOC105901894 (LOC105901894), transcript variant X2, ncRNA

Sequence ID: **ref|XR_001162457.1|** Length: 79 Number of Matches: 1
Range 1: 14 to 71

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  71  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  14
```

PREDICTED: Clupea harengus uncharacterized LOC105901894 (LOC105901894), transcript variant X1, ncRNA

Sequence ID: **ref|XR_001162456.1|** Length: 84 Number of Matches: 1
Range 1: 19 to 76

IDT Exhibit 1018.0841

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  76  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  19
```

PREDICTED: Clupea harengus uncharacterized LOC105898768 (LOC105898768), ncRNA
Sequence ID: **ref|XR_001162326.1|** Length: 96 Number of Matches: 1
Range 1: 26 to 83

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  83  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  26
```

PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X4, ncRNA
Sequence ID: **ref|XR_001162263.1|** Length: 158 Number of Matches: 1
Range 1: 37 to 94

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  37  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  94
```

PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X3, ncRNA
Sequence ID: **ref|XR_001162262.1|** Length: 177 Number of Matches: 1
Range 1: 56 to 113

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1    AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  56   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  113
```

PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X2, ncRNA
Sequence ID: **ref|XR_001162261.1|** Length: 166 Number of Matches: 1
Range 1: 45 to 102

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1    AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  45   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  102
```

PREDICTED: Clupea harengus uncharacterized LOC105897580 (LOC105897580), transcript variant X1, ncRNA
Sequence ID: **ref|XR_001162260.1|** Length: 161 Number of Matches: 1
Range 1: 40 to 97

IDT Exhibit 1018.0842

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  40  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  97
```

PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X3, ncRNA
Sequence ID: **ref|XR_001162196.1|** Length: 111 Number of Matches: 1
Range 1: 44 to 101

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1    AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  101  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  44
```

PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X2, ncRNA
Sequence ID: **ref|XR_001162195.1|** Length: 104 Number of Matches: 1
Range 1: 37 to 94

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  94  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  37
```

PREDICTED: Clupea harengus uncharacterized LOC105895935 (LOC105895935), transcript variant X1, ncRNA
Sequence ID: **ref|XR_001162194.1|** Length: 107 Number of Matches: 1
Range 1: 40 to 97

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  97  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  40
```

Dicrocoelium dendriticum genome assembly D_dendriticum_Leon_v1_0_4, scaffold DDEL_contig0063667
Sequence ID: **emb|LK829830.1|** Length: 603 Number of Matches: 1
Range 1: 540 to 597

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1    AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  597  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  540
```

Nippostrongylus brasiliensis genome assembly N_brasiliensis_RM07_v1_5_4 ,scaffold NBR_scaffold0003529
Sequence ID: **emb|LM437148.1|** Length: 19176 Number of Matches: 1
Range 1: 19117 to 19174

IDT Exhibit 1018.0843

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1      AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  19174  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  19117
```

Nippostrongylus brasiliensis genome assembly N_brasiliensis_RM07_v1_5_4 ,scaffold NBR_scaffold0003502
Sequence ID: **emb|LM437114.1|** Length: 19368 Number of Matches: 1
Range 1: 3936 to 3993

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1      AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3993   AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  3936
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000002276
Sequence ID: **emb|LN600047.1|** Length: 341394 Number of Matches: 1
Range 1: 248502 to 248559

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1      AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  248559 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  248502
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000000045
Sequence ID: **emb|LN598550.1|** Length: 947182 Number of Matches: 1
Range 1: 57307 to 57364

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1      AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  57307  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  57364
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000003581
Sequence ID: **emb|LN598496.1|** Length: 28110 Number of Matches: 1
Range 1: 22731 to 22788

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1      AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22731  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  22788
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000001750
Sequence ID: **emb|LN598313.1|** Length: 345239 Number of Matches: 1
Range 1: 3317 to 3374

IDT Exhibit 1018.0844

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Minus	

Features:

```
Query  1     AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3374  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  3317
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000001003

Sequence ID: **emb|LN598283.1|** Length: 1505843 Number of Matches: 1

Range 1: 257631 to 257688

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1       AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  257631  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  257688
```

Cyprinus carpio genome assembly common carp genome ,scaffold 000001567

Sequence ID: **emb|LN598209.1|** Length: 1073906 Number of Matches: 1

Range 1: 116306 to 116363

Score	Expect	Identities	Gaps	Strand	Frame
105 bits(116)	4e-20()	58/58(100%)	0/58(0%)	Plus/Plus	

Features:

```
Query  1       AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  58
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  116306  AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT  116363
```


IDT Exhibit 1018.0845

PALM INTRANET

Day : Tuesday
Date: 2/23/2016

Time: 15:45:15

Inventor Name Search

Enter the first few letters of the Inventor's Last Name.
Additionally, enter the first few letters of the Inventor's First name.

Last Name	First Name	
Amorese	Douglas	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1018.0846

PALM INTRANET

Day : Tuesday
Date: 2/23/2016

Time: 15:45:15

Inventor Name Search

Enter the first few letters of the Inventor's Last Name.
Additionally, enter the first few letters of the Inventor's First name.

Last Name	First Name	
Scolnick	Jonathan	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1018.0847

PALM INTRANET

Day : Tuesday
Date: 2/23/2016

Time: 15:45:15

Inventor Name Search

Enter the first few letters of the Inventor's Last Name.
Additionally, enter the first few letters of the Inventor's First name.

Last Name	First Name	
Schroeder	Ben	Search

To Go BACK Use BACK Button on Your BROWSER Tool Bar
Back to PALM | ASSIGNMENT | OASIS | Home page

IDT Exhibit 1018.0848

WEST Search History for Application 14540917

Creation Date: 2016022415:01

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
sequencing and reads near3 (duplicat$7 or redundan$8) and (barcod$8 or index$8) near3 random and (duplicat$7 or redundan$8) near3 remov$8	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
sequencing and reads near3 (duplicat$7 or redundan$8) and (barcod$8 or index$8) near3 (random$8 or degener$8)	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
sequencing and reads near3 (duplicat$7 or redundan$8) and (barcod$8 or index$8) near3 (random$8 or degener$8)	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
sequencing and duplicate near3 remov$8 and ((C12Q1/6869 | C12Q2563/179 | C12Q2535/122 | C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 | C12Q2521/319 | C12Q2521/507 | C12Q2525/15 | C12Q2525/155 | C12Q2525/161 | C12Q2525/185 | C12Q2525/191 | C12Q2525/197 | C12Q2525/313 | C12Q2537/159 | C12Q2537/165 | C12Q2549/10 | C12Q2565/514 | C12Q1/6874 | C12Q1/6876 | C12Q1/6879 | C12Q1/689 | C12Q2600/156 | C12Q2600/166 | C12N15/1065 | C12N15/1093 | C12N15/1089 | G06F19/18).CPC.)	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
(sequencing and duplicate near3 remov$8 and ((C12Q1/6869 | C12Q2563/179 | C12Q2535/122 | C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 | C12Q2521/319 | C12Q2521/507 | C12Q2525/15 | C12Q2525/155 | C12Q2525/161 | C12Q2525/185 | C12Q2525/191 | C12Q2525/197 | C12Q2525/313 | C12Q2537/159 | C12Q2537/165 | C12Q2549/10 | C12Q2565/514 | C12Q1/6874 | C12Q1/6876 | C12Q1/6879 | C12Q1/689 | C12Q2600/156 | C12Q2600/166 | C12N15/1065 | C12N15/1093 | C12N15/1089 | G06F19/18).CPC.)) and random$8 near3	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015

IDT Exhibit 1018.0849

(barcod$8 or index$8)						
sequencing and duplicate near3 remov$8 and ((C12Q1/6869 | C12Q2563/179 | C12Q2535/122 | C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 | C12Q2521/319 | C12Q2521/507 | C12Q2525/15 | C12Q2525/155 | C12Q2525/161 | C12Q2525/185 | C12Q2525/191 | C12Q2525/197 | C12Q2525/313 | C12Q2537/159 | C12Q2537/165 | C12Q2549/10 | C12Q2565/514 | C12Q1/6874 | C12Q1/6876 | C12Q1/6879 | C12Q1/689 | C12Q2600/156 | C12Q2600/166 | C12N15/1065 | C12N15/1093 | C12N15/1089 | G06F19/18).CPC.)	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
(sequencing and duplicate near3 remov$8 and ((C12Q1/6869 | C12Q2563/179 | C12Q2535/122 | C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 | C12Q2521/319 | C12Q2521/507 | C12Q2525/15 | C12Q2525/155 | C12Q2525/161 | C12Q2525/185 | C12Q2525/191 | C12Q2525/197 | C12Q2525/313 | C12Q2537/159 | C12Q2537/165 | C12Q2549/10 | C12Q2565/514 | C12Q1/6874 | C12Q1/6876 | C12Q1/6879 | C12Q1/689 | C12Q2600/156 | C12Q2600/166 | C12N15/1065 | C12N15/1093 | C12N15/1089 | G06F19/18).CPC.)) and random$8 near3 (barcod$8 or index$8)	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
sequencing and duplicate near3 remov$8 and ((C12Q1/6869 | C12Q2563/179 | C12Q2535/122 | C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 | C12Q2521/319 | C12Q2521/507 | C12Q2525/15 | C12Q2525/155 | C12Q2525/161 | C12Q2525/185 | C12Q2525/191 | C12Q2525/197 | C12Q2525/313 | C12Q2537/159 | C12Q2537/165 | C12Q2549/10 | C12Q2565/514 | C12Q1/6874 | C12Q1/6876 | C12Q1/6879 | C12Q1/689 | C12Q2600/156 | C12Q2600/166 | C12N15/1065 | C12N15/1093 | C12N15/1089 | G06F19/18).CPC.)	DWPI	n/a	ADJ	YES	ASSIGNEE	05-13-2015
20110269647	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
(Amorese.in. or Scolnick.in. or Schroeder.in. or Nugen.asn.) and sequencing and duplicate	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015

IDT Exhibit 1018.0850

(Amorese.in. or Scolnick.in. or Schroeder.in. or Nugen.asn.) and sequencing and duplicate	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
US-6251639-$.DID. OR US-5750341-$.DID. OR US-6306597-$.DID. OR US-5969119-$.DID. OR US-7244559-$.DID. OR US-7335762-$.DID. OR US-7211390-$.DID. OR US-7244567-$.DID.	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
US-7264929-$.DID. OR US-7323305-$.DID. OR US-7501245-$.DID. OR US-7491498-$.DID. OR US-7276720-$.DID. OR US-7462452-$.DID. OR US-7476504-$.DID. OR US-7405281-$.DID. OR US-7170050-$.DID. OR US-7462468-$.DID. OR US-7476503-$.DID.	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
US-7315019-$.DID. OR US-7302146-$.DID. OR US-7313318-$.DID. OR US-6449562-$.DID. OR US-6287766-$.DID. OR US-7361468-$.DID. OR US-6225109-$.DID. OR US-6110709-$.DID.	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
US-20130231253-$.DID. OR US-20060024711-$.DID. OR US-20090061439-$.DID. OR US-20080087826-$.DID. OR US-20060286566-$.DID. OR US-20060024678-$.DID. OR US-20080213770-$.DID. OR US-20080103058-$.DID. OR US-20090029385-$.DID. OR US-20090068655-$.DID.	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
US-20090024331-$.DID. OR US-20080206764-$.DID. OR US-20090026082-$.DID.	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
((US-20130231253-$.DID. OR US-20060024711-$.DID. OR US-20090061439-$.DID. OR US-20080087826-$.DID. OR US-20060286566-$.DID. OR US-20060024678-$.DID. OR US-20080213770-$.DID. OR US-20080103058-$.DID. OR US-20090029385-$.DID. OR US-20090068655-$.DID.) or (US-20090024331-$.DID. OR US-20080206764-$.DID. OR US-20090026082-$.DID.)) and sequencing and duplicate	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015

((US-20130231253-$.DID. OR US-20060024711-$.DID. OR US-20090061439-$.DID. OR US-20080087826-$.DID. OR US-20060286566-$.DID. OR US-20060024678-$.DID. OR US-20080213770-$.DID. OR US-20080103058-$.DID. OR US-20090029385-$.DID. OR US-20090068655-$.DID.) or (US-20090024331-$.DID. OR US-20080206764-$.DID. OR US-20090026082-$.DID.)) and sequencing and random near3 (barcod$8 or index$8)	PGPB	n/a	ADJ	YES	ASSIGNEE	05-13-2015
((US-6251639-$.DID. OR US-5750341-$.DID. OR US-6306597-$.DID. OR US-5969119-$.DID. OR US-7244559-$.DID. OR US-7335762-$.DID. OR US-7211390-$.DID. OR US-7244567-$.DID.) or (US-7264929-$.DID. OR US-7323305-$.DID. OR US-7501245-$.DID. OR US-7491498-$.DID. OR US-7276720-$.DID. OR US-7462452-$.DID. OR US-7476504-$.DID. OR US-7405281-$.DID. OR US-7170050-$.DID. OR US-7462468-$.DID. OR US-7476503-$.DID.) or (US-7315019-$.DID. OR US-7302146-$.DID. OR US-7313318-$.DID. OR US-6449562-$.DID. OR US-6287766-$.DID. OR US-7361468-$.DID. OR US-6225109-$.DID. OR US-6110709-$.DID.)) and sequencing and duplicate	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015
((US-6251639-$.DID. OR US-5750341-$.DID. OR US-6306597-$.DID. OR US-5969119-$.DID. OR US-7244559-$.DID. OR US-7335762-$.DID. OR US-7211390-$.DID. OR US-7244567-$.DID.) or (US-7264929-$.DID. OR US-7323305-$.DID. OR US-7501245-$.DID. OR US-7491498-$.DID. OR US-7276720-$.DID. OR US-7462452-$.DID. OR US-7476504-$.DID. OR US-7405281-$.DID. OR US-7170050-$.DID. OR US-7462468-$.DID. OR US-7476503-$.DID.) or (US-7315019-$.DID. OR US-7302146-$.DID. OR US-7313318-$.DID. OR US-6449562-$.DID. OR US-6287766-$.DID. OR	USPT	n/a	ADJ	YES	ASSIGNEE	05-13-2015

IDT Exhibit 1018.0852

US-7361468-$.DID. OR US-6225109-$.DID. OR US-6110709-$.DID.)) and sequencing and random near3 (barcod$8 or index$8)						
sequencing and reads near3 (duplicat$7 or redundan$8) and (barcod$8 or index$8) near3 (random$8 or degener$8)	USPT	10	ADJ	YES	ASSIGNEE	02-24-2016
sequencing and reads near3 (duplicat$7 or redundan$8) and (barcod$8 or index$8) near3 (random$8 or degener$8)	PGPB	33	ADJ	YES	ASSIGNEE	02-24-2016
sequencing and duplicate near3 remov$8	USPT	314	ADJ	YES	ASSIGNEE	02-24-2016
(sequencing and duplicate near3 remov$8) and duplicate near3 (barcod$7 or index$7 or tag$7)	USPT	9	ADJ	YES	ASSIGNEE	02-24-2016
sequencing and duplicate near3 read and duplicate near3 remov$8	USPT	9	ADJ	YES	ASSIGNEE	02-24-2016
sequencing and duplicate near3 read and duplicate near3 remov$8	PGPB	58	ADJ	YES	ASSIGNEE	02-24-2016
(sequencing and duplicate near3 read and duplicate near3 remov$8) and (barcod$5 or index$7 or tag$7)	PGPB	52	ADJ	YES	ASSIGNEE	02-24-2016
sequencing and duplicate near3 read and duplicate near3 remov$8	DWPI	2	ADJ	YES	ASSIGNEE	02-24-2016
WO-2014171898-A2.did.	EPAB	0	ADJ	YES	ASSIGNEE	02-24-2016

IDT Exhibit 1018.0853

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14540917	AMORESE ET AL.
	Examiner	Art Unit
	TERESA E STRZELECKA	1637

CPC- SEARCHED		
Symbol	Date	Examiner
C12Q1/6869 | C12Q2563/179 | C12Q2535/122 |C12Q2537/1376 | C12Q2563/159 | C12Q1/6806 |C12Q2521/319 | C12Q2521/507 | C12Q2525/15 |C12Q2525/155 | C12Q2525/161 | C12Q2525/185 |C12Q2525/191 | C12Q2525/197	5/13/15	/TS/

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
IDS (12/4/14; 3/21/15)	5/13/15	/TS/
WEST	5/13/15	/TS/
Google	5/13/15	/TS/
Inventor name search	5/13/15	/TS/
IDS (8/19/15)	9/21/15	/TS/
Inventor name search	2/23/16	/TS/
Google	2/23/16	/TS/
BLAST	2/23/16	/TS/
WEST	2/24/16	/TS/

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner

IDT Exhibit 1018.0854

Google+ Search Images Maps Play YouTube News Gmail More Sign in

Google next generation sequencing duplicate reads Search

All Images Videos News Shopping Maps Books

About 105,000 results

Any time
Past hour
Past 24 hours
Past week
Past month
Past year

All results
Verbatim

Next-Gen Sequencing (NGS) - thermofisher.com
www.thermofisher.com/ion-s5
Ion S5 **Next-Generation Sequencing** 15 Minute Setup, 2.5 Hour Run Times
Sequencing Reagents · Comprehensive Sequencing · Gold Standard Sanger

Next-Gen Seq Applications | PGM sequencer
AmpliSeq Panels | SOLiD 5500W Analyzer

How PCR **duplicates** arise in **next-generation sequencing**
www.cureffi.org/.../how-pcr-**duplicates**-arise-in-**next-generation-sequencing**/
Dec 11, 2012 ... This will let you have the fraction of your final **reads** that are PCR **duplicates** can be as low as 4%. Higher rates of PCR **duplicates** e.g. 30% ...

What Is Pcr **Duplicates** Of Ngs? - BioStar
https://www.biostars.org/p/15818/
Is it **duplicates** when preparing libraries, or produced during bridge amplification step of Illumina ... Or, only 1 out of 1000000 of our **reads** is really valuable?? This is ... What'S The Meaning Of "Bridge Pcr" In The **Next Generation Sequencing**.

Read **Duplicates** - BioStar
https://www.biostars.org/p/47229/
I am doing differential gene expression analysis by **next-generation sequencing**. NGS generates read **duplicates** and there are several programs available for the ... An artificial site tends to show heavy imbalances by strand, with most **reads**

[PDF] **Next-Generation Sequencing** Technologies - Navin lab
www.navinlab.com/bioinfo/**Next-Gen-Sequencing**_NAVIN.pdf
Next-Generation Sequencing ... **Generations** of **Sequencing** Technologies ..... PCR **duplicates** usually constitute about 10% of the total sequence **reads**.

Next Generation Sequencing (NGS)/Alignment - Wikibooks, open ...
https://en.wikibooks.org/.../**Next_Generation_Sequencing**.../Alignment
Next-generation sequencing generally produces short **reads** or short read pairs, ... A related error would be PCR **duplicates**, where the same read pair occurs ...
Introduction · Short reads · Alignment · Software packages

Removing **duplicates** is it really necessary? - SEOanswers
seoanswers.com/forums/showthread.php?t=6854
If you define **duplicates** as **reads** with identical sequence you may underestimate the ... of the **reads** avoids the influence of library **sequencing** error rate ... to PCR amplification bias is to reduce their **generation** in the first place.

How important are those NGS metrics? - Integrated DNA Technologies
https://www.idtdna.com/.../how-important-are-those-ngs-metrics
Jun 12, 2014 ... **Duplicate reads**, especially in paired-end sequencing, are assumed to be .... for targeted **next generation sequencing** of defined gene families..

FastUniq: A Fast De Novo **Duplicates** Removal Tool for Paired Short ...
www.ncbi.nlm.nih.gov/pmc/articles/PMC3527383/
Dec 20, 2012 ... The presence of **duplicates** introduced by PCR amplification is a major issue in paired short **reads** from **next-generation sequencing** platforms.

[PDF] Bias from removing read duplication in ultra-deep **sequencing** ...
bioinformatics.oxfordjournals.org/content/.../bioinformatics.btt771.full.pdf
Jan 2, 2014 ... read duplication is non-negligible, and removing **duplicate reads** ... **reads** from **next-generation sequencing** (NGS) experiments (Mardis, 2011).

Duplicate reads removal - OMICtools
omictools.com/**duplicate-reads**-removal-category
Duplicate read removal software tools | High-throughput ... ... amplification is a major issue in paired short **reads** from **next-generation sequencing** platforms.

Next Gen Sequencing - Power your science on Google Cloud

IDT Exhibit 1018.0855

cloud.google.com/genomics
Open, Secure & Powerful Informatics
Unlimited Storage · Pay Only For What You Use · $300 Free Trial
Cloud DNS · All Products · BigQuery · Contact Our Sales Team

Searches related to **next generation sequencing duplicate reads**

next generation sequencing **wiki**	next generation sequencing **pdf**
next generation sequencing **definition**	next generation sequencing **review**
next generation sequencing **ppt**	next generation sequencing **youtube**
next generation sequencing **tutorial**	next generation **dna** sequencing

Google
1 2 3 4 5 6 7 8 9 10 Next

Advanced search Search Help Send feedback

Google Home Advertising Programs Business Solutions Privacy Terms
About Google

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:	AMORESE, Douglas, et al.	Confirmation No.:	7076
Application No.:	14/540,917	Group Art Unit:	1637
Filed:	November 13, 2014	Examiner:	Teresa E. Strzelecka

FOR: **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

Filed *via* EFS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT/RESPONSE TO OFFICE ACTION

This paper is filed in response to the Office Action dated March 3, 2016. Please amend the above-identified patent application in the following manner:

Amendments to the Claims begin on page 2 of this paper.

Remarks/Arguments begin on page 6 of this paper.

Electronic Patent Application Fee Transmittal

Application Number:	14540917
Filing Date:	13-Nov-2014
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas AMORESE
Filer:	Jeffrey R. Bolland
Attorney Docket Number:	NUGN-001/01US

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	2806	1	90	90
		Total in USD ($)		**90**

CLAIMS:

1. (Currently Amended) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

    a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

        (i) an indexing primer binding site;

        (ii) an indexing site;

        (iii) an identifier site consisting of between 1 and 8 nucleotides; and

        (iv) a target sequence primer binding site;

    b) amplifying the adapter-nucleic acid fragment ligated products;

    c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

    d) identifying a duplicate sequencing read as the sequencing read comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads;

    wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is variable in sequence content in a plurality of adaptors.

2. (Original) The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.

3. (Original) The method of claim 1, wherein the identifier site is sequenced with the indexing site.

4. (Cancelled)

5. (Cancelled)

6. (Cancelled)

7. (Original) The method of claim 1, wherein the adaptor comprises from 5' to 3':

    (i) the indexing primer binding site;

    (ii) the indexing site;

    (iii) the identifier site; and

    (iv) the target sequence primer binding site.

8. (Cancelled)

9. (Original) The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.

10. (Previously Presented) The method of claim 9, wherein the nucleic acid fragments from each sample have the same indexing site, and each sample has a unique indexing site.

11. (Original) The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. (Original) The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. (Original) The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. (Original) The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. (Original) The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. (Original) The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. (Previously Presented) The method of claim 1, wherein the identifier site is 6 nucleotides in length.

18. (Previously Presented) The method of claim 1, wherein the identifier site is 8 nucleotides in length.

19. (Original) The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. (Original) The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

21. (Cancelled)

22. (Cancelled)

23. (Cancelled)

24. (Previously Presented) The method of claim 10, wherein the method further comprises separating fragments or sequencing reads based on indexing site.

25. (New) A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:

a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:

(i) an indexing primer binding site;

(ii) an indexing site;

(iii) an identifier site consisting of between 1 and 8 nucleotides; and

(iv) a target sequence primer binding site;

b) amplifying the adapter-nucleic acid fragment ligated products;

c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and

d) identifying a duplicate sequencing read as the sequencing read comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads;

wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides; and wherein the sequence of the identifier site is not variable in sequence content in a plurality of adaptors.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:	AMORESE, Douglas, et al.	Confirmation No.:	7076
Application No.:	14/540,917	Group Art Unit:	1637
Filed:	November 13, 2014	Examiner:	Teresa E. Strzelecka

FOR: **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

Via EFS or
Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT UNDER 37 C.F.R. §1.97(c)

In accordance with the duty of disclosure set forth in 37 C.F.R. §1.56, Applicant(s) hereby submits the following information in conformance with 37 C.F.R. §§1.97 and 1.98.

☒ Pursuant to 37 C.F.R. §1.98, a copy of each non-US patent document cited in the attached Form PTO/SB/08 is enclosed.

☐ No copies of the publications listed on the attached Form PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98(d) because the publications were previously cited by or submitted to the Office in prior Application Serial No. _____ to which the above-identified application claims priority under 35 U.S.C. §120.

☒ No copies of any U.S. patents or U.S. patent application publications listed on the attached Form PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98.

☐ Publications listed on the attached Form PTO/SB/08 were cited in a foreign search or examination report corresponding to _____ application serial no. _____ and mailed on _____

☐ Enclosed is a copy of a non-English publication(s) _____. Pursuant to §609 of the M.P.E.P., Applicant submits the attached foreign search or examination report, which cites such non-English language publication(s).

☐ Enclosed is a copy of a non-English publication(s) _____ English language publication _____ (copy enclosed) claims priority from this non-English publication.

☐ Enclosed is an explanation of non-English publication(s) _____ for which an English translation is not available.

☐ Enclosed is an English translation of non-English publication(s) _____ cited in the attached Form PTO/SB/08.

☐ Enclosed is a copy of pending patent Application Serial No. _____

This Information Disclosure Statement is filed after the period specified in 37 C.F.R. § 1.97(b), but before the mailing of any of the following:

☒ a final action under 37 C.F.R. §1.113;

☒ a notice of allowance under 37 C.F.R. §1.311; or

☒ an action that otherwise closes prosecution in this application.

In accordance with 37 C.F.R. §1.97(c) also enclosed is:

☐ Fee under 37 C.F.R. §1.17(p) in the amount of $180.00;

☒ Fee under 37 C.F.R. §1.17(p) in the amount of $90.00;

☐ Fee under 37 C.F.R. §1.17(p) in the amount of $45.00; or

☐ Statement as specified in 37 C.F.R. §1.97(e):

☐ Each item of information contained in the Information Disclosure Statement cited herein was first cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing date of the Information Disclosure Statement; or

☐ No item of information contained in the Information Disclosure Statement submitted herewith was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the undersigned, having made a reasonable inquiry, no item of information contained in the Information Disclosure Statement was known to any individual designated in 37 C.F.R. §1.56(c) more than three months prior to the filing date of the Information Disclosure Statement.

It is respectfully requested that the Examiner consider the above-noted information and return an initialed copy of the attached Form PTO/SB/08 to the undersigned.

Dated: May 31, 2016

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (202) 728-7017
Fax: (202) 842-7899

By: /Jeffrey R. Bolland/
Jeffrey R. Bolland, Ph.D.
Reg. No. 73,759

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 March 2015 (05.03.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/031691 A1

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

(21) International Application Number:
PCT/US2014/053301

(22) International Filing Date:
28 August 2014 (28.08.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/871,232	28 August 2013 (28.08.2013)	US
61/952,036	12 March 2014 (12.03.2014)	US
62/012,237	13 June 2014 (13.06.2014)	US

(71) Applicant: **CELLULAR RESEARCH, INC.** [US/US]; 3183 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors: **FAN, Christina**; 5664 Impatiens Common, Fremont, CA 94538 (US). **FODOR, Stephen, P.a.**; 1120 Parkinson Avenue, Palo Alto, CA 94301 (US). **FU, Glenn, K.**; 2771 East Cog Hill Terrace, Dublin, CA 94568 (US). **FACER, Geoffrey, Richard**; 1671 Kansas Avenue, Redwood City, CA 94061 (US). **WILHELMY, Julie**; 1726 Escalone Drive, Santa Cruz, CA 95060 (US).

(74) Agents: **ANDRES, Charles, Jr.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(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, LT, 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.

(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).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: MASSIVELY PARALLEL SINGLE CELL ANALYSIS

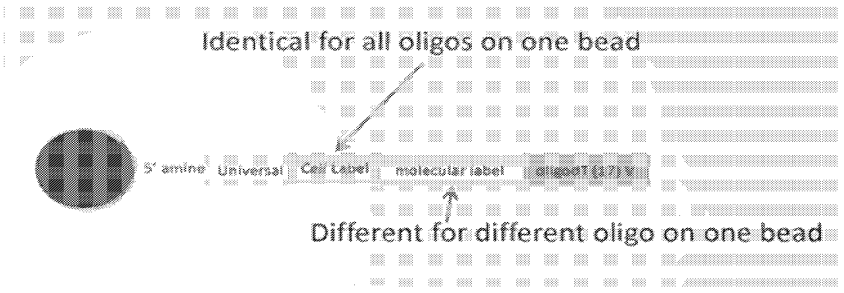


FIG. 1

(57) **Abstract**: The disclosure provides for methods, compositions, and kits for multiplex nucleic acid analysis of single cells. The methods, compositions and systems may be used for massively parallel single cell sequencing. The methods, compositions and systems may be used to analyze thousands of cells concurrently. The thousands of cells may comprise a mixed population of cells (e.g., cells of different types or subtypes, different sizes).

WO 2015/031691 A1

IDT Exhibit 1018.0867

MASSIVELY PARALLEL SINGLE CELL ANALYSIS

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/012,237, filed on June 13, 2014, U.S. Provisional Application No. 61/952,036, filed on March 12, 2014, and U.S. Provisional Application No. 61/871,232, filed on August 28, 2013, each of which applications are incorporated herein by reference in their entirety.

BACKGROUND

[0002] Multicellular masses, such as tissues and tumors, may comprise a heterogeneous cellular milieu. These complex cellular environments may often display multiple phenoytpes, which may be indicative of multiple genotypes. Distilling multicellular complexity down to single cell variability is an important facet of understanding multicellular heterogeneity. This understanding may be important in the development of therapeutic regimens to combat diseases with multiple resistance genotypes.

SUMMARY OF THE INVENTION

[0003] One aspect provided is a method, comprising obtaining a sample comprising a plurality of cells; labeling at least a portion of two or more polynucleotide molecules, complements thereof, or reaction products therefrom, from a first cell of the plurality and a second cell of the plurality with a first same cell label specific to the first cell and a second same cell label specific to the second cell; and a molecular label specific to each of the two or more polynucleotide molecules, complements thereof, or reaction products therefrom, wherein each molecular label of the two or more polynucleotide molecules, complements thereof, or reaction products therefrom, from the first cell are unique with respect to each other, and wherein each molecular label of the two or more polynucleotide molecules, complements thereof, or reaction products therefrom, from the second cell are unique with respect to each other. In some embodiments, the method further comprises sequencing the at least a portion of two or more polynucleotide molecules, complements thereof, or reaction products therefrom. In some embodiments, the method further comprises analyzing sequence data from the sequencing to identify a number of individual molecules of the polynucleotides in a specific one of the cells. In some embodiments, the cells are cancer cells. In some embodiments, the cells are infected with viral polynucleotides. In some embodiments, the cells are bacteria or fungi. In some embodiments, the sequencing comprises sequencing with

IDT Exhibit 1018.0868

read lengths of at least 100 bases. In some embodiments, the sequencing comprises sequencing with read lengths of at least 500 bases. In some embodiments, the polynucleotide molecules are mRNAs or micro RNAs, and the complements thereof and reaction products thereof are complements of and reaction products therefrom the mRNAs or micro RNAs. In some embodiments, the molecular labels are on a bead. In some embodiments, the label specific to an individual cell is on a bead. In some embodiments, the label specific to an individual cell and the molecular labels are on beads. In some embodiments, the method is performed at least in part in an emulsion. In some embodiments, the method is performed at least in part in a well or microwell of an array. In some embodiments, the presence of a polynucleotide that is associated with a disease or condition is detected. In some embodiments, the disease or condition is a cancer. In some embodiments, at least a portion of a microRNA, complement thereof, or reaction product therefrom is detected. In some embodiments, the disease or condition is a viral infection.In some embodiments, the viral infection is from an enveloped virus. In some embodiments, the viral infection is from a non-enveloped virus. In some embodiments, the virus contains viral DNA that is double stranded. In some embodiments, the virus contains viral DNA that is single stranded. In some embodiments, the virus is selected from the group consisting of a pox virus, a herpes virus, a vericella zoster virus, a cytomegalovirus, an Epstein-Barr virus, a hepadnavirus, a papovavirus, polyomavirus, and any combination thereof. In some embodiments, the first cell is from a person not having a disease or condition and the second cell is from a person having the disease or condition. In some embodiments, the persons are different. In some embodiments, the persons are the same but cells are taken at different time points. In some embodiments, the first cell is from a person having the disease or condition and the second cell is from the same person. In some embodiments, the cells in the sample comprise cells from a tissue or organ. In some embodiments, the cells in the sample comprise cells from a thymus, white blood cells, red blood cells, liver cells, spleen cells, lung cells, heart cells, brain cells, skin cells, pancreas cells, stomach cells, cells from the oral cavity, cells from the nasal cavity, colon cells, small intestine cells, kidney cells, cells from a gland, brain cells, neural cells, glial cells, eye cells, reproductive organ cells, bladder cells, gamete cells, human cells, fetal cells, amniotic cells, or any combination thereof.

[0004] One aspect provided is a solid support comprising a plurality of oligonucleotides each comprising a cellular label and a molecular label, wherein each cellular label of the plurality of oligonucleotides are the same, and each molecular label of the plurality of oligonucleotides are different; and wherein the solid support is a bead,the cellular label is specific to the solid

IDT Exhibit 1018.0869

support, the solid support, when placed at the center of a three dimensional Cartesian coordinate system, has oligonucleotides extending into at least seven of eight octants, or any combination thereof. In some embodiments, the plurality of oligonucleotides further comprises at least one of a sample label; a universal label; and a target nucleic acid binding region. In some embodiments, the solid support comprises the target nucleic acid binding region, wherein the target nucleic acid binding region comprises a sequence selected from the group consisting of a gene-specific sequence, an oligo-dT sequence, a random multimer, and any combination thereof. In some embodiments, the solid support further comprises a target nucleic acid or complement thereof. In some embodiments, the solid support comprises a plurality of target nucleic acids or complements thereof comprising from about 0.01% to about 100% of transcripts of a transcriptome of an organism or complements thereof, or from about 0.01% to about 100% of genes of a genome of an organism or complements thereof. In some embodiments, the cellular labels of the plurality of oligonucleotides comprise a first random sequence connected to a second random sequence by a first label linking sequence; and the molecular labels of the plurality of oligonucleotides comprise random sequences. In some embodiments, the solid support is selected from the group consisting of a polydimethylsiloxane (PDMS) solid support, a polystyrene solid support, a glass solid support, a polypropylene solid support, an agarose solid support, a gelatin solid support, a magnetic solid support, a pluronic solid support, and any combination thereof. In some embodiments, the plurality of oligonucleotides comprise a linker comprising a linker functional group, and the solid support comprises a solid support functional group; wherein the solid support functional group and linker functional group connect to each other. In some embodiments, the linker functional group and the solid support functional group are individually selected from the group consisting of C6, biotin, streptavidin, primary amine(s), aldehyde(s), ketone(s), and any combination thereof. In some embodiments, molecular labels of the plurality of oligonucleotides comprise at least 15 nucleotides.

[0005] One aspect provided is a kit comprising any of the solid supports described herein, and instructions for use. In some embodiments, the kit further comprises a well. In some embodiments, the well is comprised in an array. In some embodiments, the well is a microwell. In some embodiments, the kit further comprises a buffer. In some embodiments, the kit is contained in a package. In some embodiments, the package is a box. In some embodiments, the package or box has a volume of 2 cubic feet or less. In some embodiments, the package or box has a volume of 1 cubic foot or less.

IDT Exhibit 1018.0870

[0006] One aspect provided is an emulsion comprising any of the solid supports described herein.

[0007] One aspect provided is a composition comprising a well and any of the solid supports described herein.

[0008] One aspect provided is a composition comprising a cell and any of the solid supports described herein.

[0009] In some embodiments, the emulsion or composition further comprises a cell. In some embodiments, the cell is a single cell. In some embodiments, the well is a microwell. In some embodiments, the microwell has a volume ranging from about 1,000 μm^3 to about 120,000 μm^3.

[0010] One aspect provided is a method, comprising contacting a sample with any solid support disclosed herein, hybridizing a target nucleic acid from the sample to an oligonucleotide of the plurality of oligonucleotides. In some embodiments, the method further comprises amplifying the target nucleic acid or complement thereof. In some embodiments, the method further comprises sequencing the target nucleic acid or complement thereof, wherein the sequencing comprises sequencing the molecular label of the oligonucleotide to which the target nucleic acid or complement thereof is bound. In some embodiments, the method further comprises determining an amount of the target nucleic acid or complement thereof, wherein the determining comprises quantifying levels of the target nucleic acid or complement thereof; counting a number of sequences comprising the same molecular label; or a combination thereof. In some embodiments, the method does not comprise aligning any same molecular labels or any same cellular labels. In some embodiments, the amplifying comprises reverse transcribing the target nucleic acid. In some embodiments, the amplifying employs a method selected from the group consisting of: PCR, nested PCR, quantitative PCR, real time PCR, digital PCR, and any combination thereof. In some embodiments, the amplifying is performed directly on the solid support; on a template transcribed from the solid support; or a combination thereof. In some embodiments, the sample comprises a cell. In some embodiments, the cell is a single cell. In some embodiments, the contacting occurs in a well. In some embodiments, the well is a microwell and is contained in an array of microwells.

[0011] One aspect provided is a device, comprising a plurality of microwells, wherein each microwell of the plurality of microwells has a volume ranging from about 1,000 μm^3 to about 120,000 μm^3. In some embodiments, each microwell of the plurality of microwells has a volume of about 20,000 μm^3. In some embodiments, the plurality of microwells comprises

from about 96 to about 200,000 microwells. In some embodiments, the microwells are comprised in a layer of a material. In some embodiments, at least about 10 % of the microwells further comprise a cell. In some embodiments, the device further comprises any of the solid supports described herein.

[0012] One aspect provided is an apparatus comprising any of the devices described herein, and a liquid handler. In some embodiments, the liquid handler delivers liquid to the plurality of microwells in about one second. In some embodiments, the liquid handler delivers liquid to the plurality of microwells from a single input port. In some embodiments, the apparatus further comprises a magnet. In some embodiments, the apparatus further comprises at least one of: an inlet port, an outlet port, a pump, a valve, a vent, a reservoir, a sample collection chamber, a temperature control apparatus, or any combination thereof. In some embodiments, the apparatus comprises the sample collection chamber, wherein the sample collection chamber is removable from the apparatus. In some embodiments, the apparatus further comprises an optical imager. In some embodiments, the optical imager produces an output signal which is used to control the liquid handler. In some embodiments, the apparatus further comprises a thermal cycling mechanism configured to perform a polymerase chain reaction (PCR) amplification of oligonucleotides.

[0013] One aspect provided is a method of producing a clinical diagnostic test result, comprising producing the clinical diagnostic test result with any device or apparatus described herein; any solid support described herein; any method described herein; or any combination thereof. In some embodiments, the clinical diagnostic test result is transmitted via a communication medium.

[0014] One aspect provided is a method of making any of the solid supports described herein, comprising attaching to a solid support: a first polynucleotide comprising a first portion of the cellular label, and a first linker; andcontacting a second polynucleotide comprising a second portion of the cellular label, a sequence complementary to the first liker, and the molecular label. In some embodiments, the third polynucleotide further comprises a target nucleic acid binding region.

[0015] In some embodiments, an emulsion, microwell, or well contains only one cell. In some embodiments, from 1 to 2,000,000 emulsions, microwells, or wells each contain only one cell. In some embodiments, the method comprises distributing at most one cell into each emulsion, microwell, or well. In some embodiments, a single solid support and a single cell are distributed to an emulsion, microwell, or well. In some embodiments, from 1 to 2,000,000 emulsions, microwells, or wells each have distributed thereto one cell and one solid support.

IDT Exhibit 1018.0872

In some embodiments, the method comprises distributing at most one solid support per emulsion, microwell, or well. In some embodiments, the method comprises distributing one solid support and one cell to each of from 1 to 2,000,000 microwells, emulsions, or wells. In some embodiments, cell distribution is random or non-random. In some embodiments, cell distribution is stochastic. In some embodiments, a cell is distributed by a cell sorter. In some embodiments, a cell is distributed by contacting one or more wells, microwells, or emulsions with a dilute solution of cells diluted so that at most one cell is distributed to the one or more wells, microwells, or emulsions.

[0016] In some embodiments, the target specific regions, target specific regions of the plurality of oligonucleotides, or the target specific region of the two or more polynucleotide molecules, comprise sequences complementary to two or more targets of a target panel. In some embodiments, the two or more targets of the target panel are biomarkers. In some embodiments, the biomarkers are biomarkers for a disease or condition. In some embodiments, the disease or condition is a cancer, an infection, a viral infection, an inflammatory disease, a neurodegenerative disease, a fungal disease, a bacterial infection, or any combination thereof. In some embodiments, the panel comprises from: 2-50,000, 2-40,000, 2-30,000, 2-20,000, 2-10,000, 2-9000, 2-8,000, 2-7,000, 2-6,000, 2-5,000, 2-1,000, 2-800, 2-700, 2-600, 2-500, 2-400, 2-300, 2-200, 2-100, 2-75, 2-50, 2-40, 2-30, 2-20, 2-10, or 2-5 biomarkers.

INCORPORATION BY REFERENCE

[0017] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0019] **FIG. 1** depicts an exemplary solid support conjugated with an exemplary oligonucleotide.

IDT Exhibit 1018.0873

[0020] FIG. 2A-C depicts an exemplary workflow for synthesizing oligonucleotide coupled beads using split-pool synthesis.

[0021] FIG. 3 depicts an exemplary oligonucleotide coupled bead.

[0022] FIG. 4 illustrates an exemplary embodiment of a microwell array.

[0023] FIG. 5 depicts an exemplary distribution of solid supports in a microwell array.

[0024] FIG. 6A-C show exemplary distribution cells onto microwell arrays. FIG. 6A shows the distribution of K562 cells (large cell size). FIG. 6B shows the distribution of Ramos cells (small cell size). FIG. 6C shows the distribution of Ramos cells and oligonucleotide coupled beads onto microwell arrays, with solid arrows pointing to the Ramos cells and dashed arrows pointing to the oligonucleotide coupled beads.

[0025] FIG. 7 shows exemplary statistics of the microwell volume, solid support volume, and amount of biological material obtained from lysis.

[0026] FIG. 8A-C illustrates an exemplary embodiment of bead cap sealing. FIG. 8A-B show images of a microarray well with cells and oligonucleotide beads distributed into wells of a microarray well and with larger sephadex beads used to seal the wells. Dotted arrows point to the cells, dashed arrows point to the oligonucleotide coupled beads and the solid arrows point to the sephadex beads. FIG. 8C depicts a schematic of the cell and oligonucleotide bead (*e.g.*, oligobead) deposited within a well with a sephadex bead used to seal the well.

[0027] FIG. 9 depicts a bar graph comparing amplification efficiency of GAPDH and RPL19 amplified from microwells and tubes. The grey bars represent data from the microwell. The white bars represent data from the tube.

[0028] FIG. 10 depicts an agarose gel comparing amplification specificity of three different genes directly on a solid support.

[0029] FIG. 11A-I show graphical representations of the sequencing results.

[0030] FIG. 12A-C show a histogram of the sequencing results for the K562-only sample, Ramos-only sample, and K562 + Ramos mixture sample, respectively.

[0031] FIG. 12D-E shows a graph of the copy number for genes listed in Table 3 for the Ramos-only cell sample and K562-only cell sample, respectively.

[0032] FIG. 12F-I show the copy number for individual genes.

[0033] FIG. 12J-M show graphs of the number of unique molecules per gene (y-axis) for the beads with the 100 unique barcode combinations.

[0034] FIG. 12N-O show enlarged graphs of two beads that depict the general pattern of gene expression profiles for the two cell types.

IDT Exhibit 1018.0874

[0035] FIG. 12P shows a scatter plot of results based on principal component analysis of gene expression profile of 768 beads with > 30 molecules per bead from the K562+Ramos mixture sample.

[0036] FIG. 12Q-R show histograms of the copy number per amplicon per bead for the K562-like cells (beads on the left of the first principal component based on FIG. 12P) and Ramos-like cells (beads on the right of the first principal component based on FIG. 12P), respectively.

[0037] FIG. 12S-T show the copy number per bead or single cell of the individual genes for the K562-like cells (beads on the left of the first principal component based on FIG. 12P) and Ramos-like cells (beads on the right of the first principal component based on FIG. 12P), respectively.

[0038] FIG. 13A depicts general gene expression patterns for the mouse and Ramos cells.

[0039] FIG. 13B-C show scatter plots of results based on principal component analysis of gene expression profile of the high density sample and low density sample, respectively.

[0040] FIG. 13D-E depict graphs of the read per barcode (bc) combination (y-axis) versus the unique barcode combination, sorted by the total number of molecules per bc combination (x-axis) for Ramos-like cells and mouse-like cells from the high density sample, respectively.

[0041] FIG. 13F-G depict graphs of the number of molecules per barcode (bc) combination (y-axis) versus the unique barcode combination, sorted by the total number of molecules per bc combination (x-axis) for Ramos-like cells and mouse-like cells from the high density sample, respectively.

[0042] FIG. 13H-I depict graphs of the read per barcode (bc) combination (y-axis) versus the unique barcode combination, sorted by the total number of molecules per barcode combination (x-axis) for Ramos-like cells and mouse-like cells from the low density sample, respectively.

[0043] FIG. 13J-K depict graphs of the number of molecules per barcode combination (y-axis) versus the unique barcode combination, sorted by the total number of molecules per barcode combination (x-axis) for Ramos-like cells and mouse-like cells from the low density sample, respectively.

[0044] FIG. 14 shows a graph depicting the genes on the X-axis and the log10 of the number of reads.

[0045] FIG. 15A shows a graph of the distribution of genes detected per three-part cell label (*e.g.*, cell barcode). **FIG. 15B** shows a graph of the distribution of unique molecules detected per bead (expressing the gene panel).

IDT Exhibit 1018.0875

[0046] FIG. 16 depicts the cell clusters based on the genes associated with a cell barcode.

[0047] FIG. 17A-D show the analysis of monocyte specific markers. **FIG. 17E** shows the cell cluster depicted in FIG. 16.

[0048] FIG. 18A-B show the analysis of the T cell specific markers. **FIG. 18C** shows the cell cluster depicted in FIG. 16.

[0049] FIG. 19A-B show the analysis of the CD8+ T cell specific markers. **FIG. 19C** shows the cell cluster depicted in FIG. 16.

[0050] FIG. 20A shows the analysis of CD4+ T cell specific markers. **FIG. 20B** shows the cell cluster depicted in FIG. 16.

[0051] FIG. 21A-D show the analysis of Natural Killer (NK) cell specific markers. **FIG. 20E** shows the cell cluster depicted in FIG. 16.

[0052] FIG. 22A-E show the analysis of B cell specific markers. **FIG. 22F** shows the cell cluster depicted in FIG. 16.

[0053] FIG. 23A-F show the analysis of Toll-like receptors. Toll-like receptors are mainly expressed by monocytes and some B cells. FIG. 23G shows the cell cluster depicted in FIG. 16.

[0054] FIG. 24 depicts a graph of the genes versus the log10 of the number of reads.

[0055] FIG. 25A-D shows graphs of the molecular barcode versus the number of reads or log10 of the number of reads for two genes.

[0056] FIG. 26A shows a graph of the number of genes in the panel expressed per cell barcode versus the number of unique cell barcodes/single cell. **FIG. 26B** shows a histogram of the number of unique molecules detected per bead versus frequency of the number of cells per unique cell barcode carrying a given number of molecules. **FIG. 26C** shows a histogram of the number of unique GAPDH molecules detected per bead versus frequency of the number of cells/unique cell barcode carrying a given number of molecules.

[0057] FIG. 27 shows a scatterplot of the 856 cells.

[0058] FIG. 28 shows a heat map of expression of the top 100 (in terms of the total number of molecules detected).

[0059] FIG. 29 shows a workflow for Example 12.

[0060] FIG. 30 shows a workflow for Example 13.

[0061] FIG. 31A-C. Clustering of single cells in controlled mixtures containing two distinct cell types. A. Clustering of a 1:1 mixture of K562 and Ramos cells by principal component analysis of the expression of 12 genes. The biplot shows two distinct clusters, with one cluster expressing Ramos specific genes and the other expressing K562 specific genes. B.

IDT Exhibit 1018.0876

Principal component analysis of a mixture containing a small percentage of Ramos cells in a background of primary B cells from a healthy individual using a panel of 111 genes. The color of each data point indicates the total number of unique transcript molecules detected across the entire gene panel. A set of 18 cells (circled) out of 1198 cells displays a distinct gene expression profile and with much higher transcription levels. C. Heatmap showing expression level of each gene in the top 100 cells in the sample of FIG. 31B, ranked by the total number of transcript molecules detected in the gene panel. Genes are ordered via hierarchical clustering in terms of correlation. The top 18 cells, indicated by the horizontal red bar, expressed preferentially a set of genes known to be associated with follicular lymphoma, as indicated by the vertical red bar.

[0062] FIG. 31D. PCA analysis of primary B cells with spiked in Ramos cells. Color of each data point (single cell) indicates the log of the number of transcript molecules each cell carries for the particular gene. Top 7 rows: Genes that are preferentially expressed by the subset of 18 cells that are likely Ramos cells. First row genes (from left to right) include GAPDH, TCL1A, MKI67 and BCL6. Second row genes (from left to right) include MYC, CCND3, CD81 and GNAI2. Third row of genes (from left to right) include IGBP1, CD20, BLNK and DOCK8. Fourth row of genes (from left to right) include IRF4, CD22, IGHM and AURKB. Fifth row of genes (from left to right) include CD38, CD10, LEF1 and AICDA. Sixth row of genes (from left to right) include CD40, CD27, IL4R and PRKCD. Seventh row of genes (from left to right) include RGS1, MCL1, CD79a and HLA-DRA. Last row: Genes that are expressed preferentially by a subset of primary B cells but not especially enriched in those 18 cells. Genes in the last row (from left to right) include IL6, CD23a, CCR7 and CXCR5.

[0063] FIG. 32 Expression of GAPDH. Color indicates natural log of the number of unique transcript molecules observed per cell.

[0064] FIG. 33A-F shows the principal component analysis (PCA) for monocyte associated genes. FIG. 33A shows the PCA for CD16. FIG. 33B shows the PCA for CCRvarA. FIG. 33C shows the PCA for CD14. FIG. 33D shows the PCA for S100A12. FIG. 33E shows the PCA for CD209. FIG. 33F shows the PCA for IFNGR1.

[0065] FIG. 34A-B shows the principal component analysis (PCA) for pan-T cell markers (CD3). FIG. 34A shows the PCA for CD3D and FIG. 34B shows the PCA for CD3E.

[0066] FIG. 35A-E shows the principal component analysis (PCA) for CD8 T cell associated genes. FIG. 35A shows the PCA for CD8A. FIG. 35B shows the PCA for EOMES. FIG. 35C

IDT Exhibit 1018.0877

shows the PCA for CD8B. FIG. 35D shows the PCA for PRF1. FIG. 35E shows the PCA for RUNX3.

[0067] FIG. 36A-C shows the principal component analysis (PCA) for CD4 T cell associated genes. FIG. 36A shows the PCA for CD4. FIG. 36B shows the PCA for CCR7. FIG. 36C shows the PCA for CD62L.

[0068] FIG. 37A-F shows the principal component analysis (PCA) for B cell associated genes. FIG. 37A shows the PCA for CD20. FIG. 37B shows the PCA for IGHD. FIG. 37C shows the PCA for PAX5. FIG. 37D shows the PCA for TCL1A. FIG. 37E shows the PCA for IGHM. FIG. 37F shows the PCA for CD24.

[0069] FIG. 38A-C shows the principal component analysis (PCA) for Natural Killer cell associated genes. FIG. 38A shows the PCA for KIR2DS5. FIG. 38B shows the PCA for CD16. FIG. 38C shows the PCA for CD62L.

[0070] FIG. 39 Simultaneous identification of major cell types in a human PBMC sample (632 cells) by PCA analysis of 81 genes assayed by CytoSeq Cells with highly correlated expression profile are coded with similar color.

[0071] FIG. 40A-B Correlation analysis of single cell gene expression profile of PBMC sample. 40A. A matrix showing the pairwise correlation coefficient across 632 cells in the sample. The cells are ordered such that those with highly correlated gene expression profile are grouped together. 40B. Heatmap showing the expression of each gene by each cell. The cells (columns) are ordered in the same manner as the correlation matrix above. The genes (rows) are ordered such that genes that share highly similar expression pattern across the cells are grouped together. The cell type of each cluster of cells may be identified by the group of genes the cells co-expressed. Within each major cell cluster, there is substantial degree of heterogeneity in terms of gene expression.

[0072] FIG. 41 data represents that of 731 cells from a replicate experiment of PBMC sample from the same donor. Cells with similar gene expression profile (based on hierarchical clustering using correlation coefficient) are plotted with similar color.

[0073] FIG. 42 shows a heat map demonstrating the correlation in gene expression profile between genes.

[0074] FIG. 43 Description of CytoSeq. A. Experimental procedure for CytoSeq. B. Structure of oligonucleotides attached to beads.

[0075] FIG. 44 Dissecting sub-populations of CD3+ T cells. A. PCA of Donor 1 unstimulated sample reveals two major branches of cells. The expression level (log of unique transcript molecule) of a particular gene within each cell is indicated with color. Helper T cell

IDT Exhibit 1018.0878

associated cytokine and effector genes are enriched in cells in the lower branch, while cytotoxic T cell associated genes are enriched in the upper branch. Shown here are representative genes. First row shows helper T cell related genes and include (from left to right) CD4, SELL and CCR7. Second row shows cytotoxic T cell related genes and include (from left to right) CD8A, NKG2D and EOMES. B. PCA of Donor 1 anti-CD3/anti-CD28 stimulated sample showing enrichment of expression of indicated genes to one of the two main branches representing helper and cytotoxic T cells. These genes are present at low amounts in the unstimulated sample. First two rows show genes that are known to be associated with activated T cells and include (from left to right) in the first row IRF4, CD69 and MYC and in the second row GAPDH, TNF and IFNG. The third row shows genes that are known to be associated with activated helper T cells and include (from left to right) IL2, LTA and CD40LG. The fourth row shows genes that are known to be associated with activated cytotoxic T cells and include (from left to right) CCL4, CCL3 and GZMB. C. Number of cells that contribute to the overall expression level of genes that exhibit large fold-changes when comparing stimulated over unstimulated samples in aggregate data. For several cytokines (red arrows), the contribution from only a small number of cells is responsible for large overall gene expression change in the entire population.

[0076] FIG. 45. PCA plots of T cell samples that have undergone stimulation with anti-CD28/anti-CD3 beads in the two donors, and the corresponding unstimulated samples, with emphasis on the expression of genes that clearly show preferential expression in either helper or cytotoxic subsets in the unstimulated samples. The color of each data point (single cell) indicates log(number of unique transcript molecule) per cell for the indicated gene. For each pair of stimulated and unstimulated graphs in each donor, the color range is adjusted to be the same. A. Genes that are known to be associated with both helper and cytotoxic T cells. B. Genes that are known to be associated with cytotoxic T cells. C. Genes that are known to be associated with helper T cells.

[0077] FIG. 46A-D PCA plots of T cell samples that have undergone stimulation with anti-CD28/anti-CD3 beads in the two donors, and the corresponding unstimulated samples, with emphasis on the expression of genes that are expressed in the stimulated samples but at low or undetectable level in the unstimulated samples. The color of each data point (single cell) indicates log(number of unique transcript molecule) per cell for the indicated gene. For each pair of stimulated and unstimulated graphs in each donor, the color range is adjusted to be the same. 46A and 46D. Genes that are expressed by both branches of cells upon activation. 46B. Genes that are expressed preferentially by cells in the upper branch upon activation. These

IDT Exhibit 1018.0879

genes are known to be associated with activated cytotoxic T cells. 46C. Genes that are expressed preferentially by cells in the lower branch upon activation. These genes are known to be associated with activated helper T cells.

[0078] FIG. 47 Clustering of data from Donor 1's unstimulated CD3+ T cells shows separations of CD4 and CD8 cells, as well as a group of cells that express Granzyme K and Granzyme A but little CD8. Top: Heatmap showing correlation between each pair of cells. Cells that are highly correlated are grouped together. Bottom: Heatmap showing the level of expression of each gene of each cell. Cells and genes are ordered via bidirectional hierarchical clustering.

[0079] FIG. 48. Similar to FIG. 47, but showing data from anti-CD3/anti-CD28 stimulated CD3+ T cell sample of Donor 1. Top: Heatmap showing correlation between each pair of cells. Cells that are highly correlated are grouped together. Bottom: Heatmap showing the level of expression of each gene of each cell. Cells and genes are ordered via bidirectional hierarchical clustering.

[0080] FIG. 49A-C In donor 1, large overall fold change was observed for various cytokines in the antiCD28/antiCD3 stimulated sample, as compared to the unstimulated one. A-B: The large fold changes of these cytokines were mostly contributed by only a few single cells (dots that are enclosed with squares or circles). A number of these cytokines were contributed by the same small number of cells. C: The co-expression patterns of these cytokines coincide with the signature cytokine combination for the Th2 and Th17 subsets of helper T cells.

[0081] FIG. 50A-B. Dissecting sub-populations of CD8+ T cells. A. Clustering of CytoSeq data defines two major groups of CD8+ cells – one group expresses genes shared by central memory/naïve cells, and the other group expresses genes shared by effector memory/effector cells. Shown here is data of Donor 2's unstimulated sample. Top: Heatmap showing correlation between each pair of cells. Bottom: Heatmap showing the level of expression of each gene in each cell. Cells and genes are ordered via bidirectional hierarchical clustering. B. Identification of rare antigen specific T cell by expression of gamma interferon (IFNG) in CD8+ T cells from two donors after stimulation with CMV peptide pool. Each cell is plotted on the 2D principal component space. Cells expressing IFNG (circled) are usually among those with the most total detected transcripts in the panel (indicated by the color). In donor 2, the top expressing cell (square) does not produce IFNG but expresses cytokines IL6 and IL1B. Number next to each circle indicates the rank in descending order the number of total unique transcript molecules detected for that cell.

IDT Exhibit 1018.0880

[0082] FIG. 51. Similar to Figure 50A except the data here represents that of Donor 2 CMV stimulated sample. A. Clustering of CytoSeq data defines two major groups of CD8+ cells – one group expresses genes shared by central memory/naïve cells, and the other group expresses genes shared by effector memory/effector cells. Shown here is data of Donor 2's unstimulated sample. Top: Heatmap showing correlation between each pair of cells. Bottom: Heatmap showing the level of expression of each gene in each cell. Cells and genes are ordered via bidirectional hierarchical clustering.

[0083] FIG. 52. Data are plotted in principal component space. Color indicates log(number of unique transcript molecule detected) for the particular gene. A. Genes that appear to be expressed by larger proportion of cells upon stimulation by CMV peptide pool. B. Genes that are enriched in one branch of cells. These genes are also known to be associated with naïve and central memory CD8+ T cells. C. Genes that are enriched in the other branch of cells. These genes are known to be associated with effector and effector memory CD8+ T cells. D. Granzyme K expressing cells occupy a region between the naïve/central memory and effector/effector memory cells on the PC space. E. HLA-DRA expressing cells constitute a special subset. F. Genes that are expressed in both branches of cells.

[0084] FIG. 53. Same as FIG. 50B, except the data represents those of the unstimulated controls. None of the cells in Donor 1's sample expressed IFNG, while one cell in Donor 2's sample expressed IFNG yet with overall low expression across the entire gene panel (rank 1069). Color scale is adjusted to match that of the respective graph for the stimulated sample.

[0085] FIG. 54. Heatmaps showing the heterogeneous expression of the gene panel in cells that express gamma interferon (IFNG) in CMV stimulated CD8+ T cells of Donors 1 and 2. Also shown is the cell that carries most total transcripts detected in Donor 2. This particular cell does not express IFNG but expresses strongly IL6, IL1B and CCL4. The cells and genes are ordered by bidirectional hierarchical clustering based on correlation. Cell ID refers to the rank in total number of detected transcripts of the gene panel, and are indicated in the PCA plots in FIG. 50.

[0086] FIG. 55. Amplification scheme. The first PCR amplifies molecules attached to the bead using a gene specific primer and a primer against the universal Illumina sequencing primer 1 sequence. The second PCR amplifies the first PCR products using a nested gene specific primer flanked by Illumina sequencing primer 2 sequence, and a primer against the universal Illumina sequencing primer 1 sequence. The third PCR adds P5 and P7 and sample index to turn PCR products into Illumina sequencing library. 150bp x 2 sequencing reveals

IDT Exhibit 1018.0881

the cell label and molecule label on read 1, the gene on read 2, and the sample index on index 1 read.

[0087] **FIG. 56** depicts a schematic of a workflow for analyzing molecules from a sample.

[0088] **FIG. 57** depicts a schematic of a workflow for analyzing molecules from a sample.

[0089] **FIG. 58A-B** depict agarose gels of PCR products.

[0090] **FIG. 59** depicts a plot of sequencing reads for a plurality of genes.

[0091] **FIG. 60A-D** depicts plots of the reads observed per label detected (RPLD) for Lys, Phe, Thr, and Dap spike-in controls, respectively. FIG. 60E depicts a plot of Reads versus Input.

[0092] **FIG. 61** depicts a plot of the reads observed per label detected (RPLD) for various genes.

[0093] **FIG. 62** depicts a plot of the reads observed per label detected (RPLD) for various genes.

[0094] **FIG. 63** depicts a plot of total reads (labels) versus rpld for various genes.

[0095] **FIG. 64** depicts a plot of RPKM for undetected genes.

[0096] **FIG. 65** depicts a schematic for the synthesis of molecular barcodes.

[0097] **FIG. 66A-C** depict schematics for the synthesis of molecular barcodes.

[0098] **FIG. 67** shows a schematic of a workflow for stochastically labeling nucleic acids.

[0099] **FIG. 68** is a schematic of a workflow for stochastically labeling nucleic acids.

[00100] **FIG. 69** illustrates a mechanical fixture within which microwell array substrates may be clamped, thereby forming a reaction chamber or well into which samples and reagents may be pipetted for performing multiplexed, single cell stochastic labeling / molecular indexing experiments. Upper: exploded view showing the upper and lower parts of the fixture and an elastomeric gasket for forming a leak-proof seal with the microwell array substrate. Lower: exploded side-view of the fixture.

[00101] **FIG. 70** illustrates a mechanical fixture which creates two reaction chambers or wells when a microwell array substrate is clamped within the fixture.

[00102] **FIG. 71** illustrates two examples of elastomeric (*e.g.*, polydimethylsiloxane) gaskets for use with the mechanical fixtures illustrated in FIGS. 69 and 70. The elastomeric gaskets provide for a leak-proof seal with the microwell array substrate to create a reagent well around the microwell array. The gaskets may contain one (upper), two (lower), or more openings for creating reagent wells.

[00103] **FIG. 72** depicts one embodiment of a cartridge within which a microwell array is packaged. Left: An exploded view of the cartridge illustrating (from bottom to top) the

microwell array substrate, a gasket that defines the flow cell or array chamber, a reagent and/or waste reservoir component for defining compartments to contain pre-loaded assay reagents or store spent reagents, and a cover for sealing the reagent and waste reservoirs and defining the sample inlet and outlet ports. Right: An assembled view of one embodiment of the cartridge design illustrating relief for bringing an external magnet into close proximity with the microwell array.

[00104] FIG. 73 depicts one embodiment of a cartridge designed to include onboard assay reagents with the packaged microwell array.

[00105] FIG. 74 provides a schematic illustration of an instrument system for performing multiplexed, single cell stochastic labeling / molecular indexing assay. The instrument system may provide a variety of control and analysis capabilities, and may be packaged as individual modules or as a fully integrated system. Microwell arrays may be integrated with flow cells that are either a fixed component of the system or are removable, or may be packaged within removable cartridges that further comprise pre-loaded assay reagent reservoirs and other functionality.

[00106] FIG. 75 illustrates one embodiment of the process steps to be performed by an automated system for performing multiplexed, single cell stochastic labeling / molecular indexing assays.

[00107] FIG. 76 illustrates one embodiment of a computer system or processor for providing instrument control and data analysis capabilities for the assay system presently disclosed.

[00108] FIG. 77 shows a block diagram illustrating one example of a computer system architecture that can be used in connection with example embodiments of the assay systems of the present disclosure.

[00109] FIG. 78 depicts a diagram showing a network with a plurality of computer systems, cell phones, personal data assistants, and Network Attached Storage (NAS), that can be used with example embodiments of the assay systems of the present disclosure.

[00110] FIG. 79 depicts a block diagram of a multiprocessor computer system that can be used with example embodiments of the assay systems of the present disclosure.

[00111] FIG. 80 depicts a diagram of analysis of a test sample and communication of test result obtained from the test sample via a communication media.

IDT Exhibit 1018.0883

DETAILED DESCRIPTION

[00112] Disclosed herein are methods, kits, and compositions for analyzing molecules in a plurality of samples. Generally, the methods, kits, and compositions comprise (a) stochastically labeling molecules in two or more samples with molecular barcodes to produce labeled molecules; and (b) detecting the labeled molecules. The molecular barcodes may comprise one or more target specific regions, label regions, sample index regions, universal PCR regions, adaptors, linkers, or a combination thereof. The labeled molecules may comprise a) a molecule region; b) a sample index region; and c) a label region. The molecule region may comprise at least a portion of the molecule from the molecular barcode was originally attached to. The molecule region may comprise a fragment of the molecule from the molecular barcode was originally attached to. The sample index region may be used to determine the source of the molecule region. The sample index region may be used to determine from which sample the molecule region originated from. The sample index region may be used to differentiate molecule regions from two or more different samples. The label region may be used to confer a unique identity to identical molecule regions originating from the same source. The label region may be used to confer a unique identity to identical molecule regions originating from the same sample.

[00113] The method for analyzing molecules in a plurality of samples may comprise: a) producing a plurality of sample-tagged nucleic acids by: i) contacting a first sample comprising a plurality of nucleic acids with a plurality of first sample tags to produce a plurality of first sample-tagged nucleic acids; and ii) contacting a second sample comprising a plurality of nucleic acids with a plurality of second sample tags to produce a plurality of second sample-tagged nucleic acids, wherein the plurality of second sample tags are different from the first sample tags; b) contacting the plurality of sample-tagged nucleic acids with a plurality of molecular identifier labels to produce a plurality of labeled nucleic acids; and c) detecting at least a portion of the labeled nucleic acids, thereby determining a count of a plurality of nucleic acids in a plurality of samples. The plurality of samples may comprise a single cell.

[00114] Alternatively, the method for analyzing molecules in a plurality of samples may comprise: a) producing a plurality of labeled nucleic acids comprising: i) contacting a first sample with a first plurality of sample tags, wherein the first plurality of sample tags comprises identical nucleic acid sequences; ii) contacting the first sample with a first plurality of molecular identifier labels may comprise different nucleic acid sequences, wherein contacting the first sample with the first plurality of sample tags or first plurality of molecular

IDT Exhibit 1018.0884

identifier labels occurs simultaneously or sequentially to produce a plurality of first-labeled nucleic acids; iii) contacting a second sample with a second plurality of sample tags, wherein the second plurality of sample tags may comprise identical nucleic acid sequences; iv) contacting the second sample with a second plurality of molecular identifier labels may comprise different nucleic acid sequences, wherein contacting the second sample with the second plurality of sample tags or second plurality of molecular identifier labels occurs simultaneously or sequentially to produce a plurality of second-labeled nucleic acids, wherein the plurality of labeled nucleic acids may comprise the plurality of first-labeled nucleic acids and the second-labeled nucleic acids; and b) determining a number of different labeled nucleic acids, thereby determining a count of a plurality of nucleic acids in a plurality of samples.

[00115] The method for analyzing molecules in a plurality of samples may comprise: a) contacting a plurality of samples may comprise two or more different nucleic acids with a plurality of sample tags and a plurality of molecular identifier labels to produce a plurality of labeled nucleic acids, wherein: i) the plurality of labeled nucleic acids may comprise two or more nucleic acids attached to two or more sample tags and two or more molecular identifier labels; ii) the sample tags attached to nucleic acids from a first sample of the plurality of samples are different from the sample tags attached to nucleic acid molecules from a second sample of the plurality of samples; and iii) two or more identical nucleic acids in the same sample are attached to two or more different molecular identifier labels; and b) detecting at least a portion of the labeled nucleic acids, thereby determining a count of two or more different nucleic acids in the plurality of samples.

[00116] FIG. 56 depicts an exemplary workflow for the quantification of RNA molecules in a sample. As shown in Step 1 of FIG. 56, RNA molecules (110) may be reverse transcribed to produce cDNA molecules (105) by the stochastic hybridization of a set of molecular identifier labels (115) to the polyA tail region of the RNA molecules. The molecular identifier labels (115) may comprise an oligodT region (120), label region (125), and universal PCR region (130). The set of molecular identifier labels may contain 960 different types of label regions. As shown in Step 2 of FIG. 56, the labeled cDNA molecules (170) may be purified to remove excess molecular identifier labels (115). Purification may comprise Ampure bead purification. As shown in Step 3 of FIG. 56, the labeled cDNA molecules (170) may be amplified to produce a labeled amplicon (180). Amplification may comprise multiplex PCR amplification. Amplification may comprise a multiplex PCR amplification with 96 multiplex primers in a single reaction volume. Amplification may

IDT Exhibit 1018.0885

comprise a custom primer (135) and a universal primer (140). The custom primer (135) may hybridize to a region within the cDNA (105) portion of the labeled cDNA molecule (170). The universal primer (140) may hybridize to the universal PCR region (130) of the labeled cDNA molecule (170). As shown in Step 4, the labeled amplicons (180) may be further amplified by nested PCR. The nested PCR may comprise multiplex PCR with 96 multiplex primers in a single reaction volume. Nested PCR may comprise a custom primer (145) and a universal primer (140). The custom primer (135) may hybridize to a region within the cDNA (105) portion of the labeled amplicon (180). The universal primer (140) may hybridize to the universal PCR region (130) of the labeled amplicon (180). As shown in Step 5, one or more adaptors (150, 155) may be attached to the labeled amplicon (180) to produce an adaptor-labeled amplicon (190). The one or more adaptors may be attached to the labeled amplicon (180) via ligation. As shown in Step 6, the one or more adaptors (150, 155) may be used to conduct one or more additional assays on the adaptor-labeled amplicon (190). The one or more adaptors (150, 155) may be hybridized to one or more primers (160, 165). The one or more primers (160, 165) be PCR amplification primers. The one or more primers (160, 165) may be sequencing primers. The one or more adaptors (150, 155) may be used for further amplification of the adaptor-labeled amplicons. The one or more adaptors (150, 155) may be used for sequencing the adaptor-labeled amplicon.

[00117] FIG. 57 depicts an exemplary schematic of a workflow for analyzing nucleic acids from two or more samples. As shown in FIG. 57, a method for analyzing nucleic acids from two or more samples may comprise selecting two or more genes for analysis and designing custom primers based on the selected genes (210). The method may further comprise supplementing one or more samples comprising nucleic acids (*e.g.*, RNA) with one or more spike-in controls (220). The nucleic acids in the sample may be amplified by multiplex RT-PCR (230) with molecular barcodes (or sample tags or molecular identifier labels) and the custom primers to produce labeled amplicons. The labeled amplicons may further treated with one or more sequencing adaptors to produce adaptor labeled amplicons (240). The adaptor labeled amplicons can be analyzed (250). As shown in FIG. 57, analysis of the labeled amplicons (250) may comprise one or more of (1) detection of a universal PCR primer seq, polyA and/or molecular barcode (or sample tag, molecular identifier label); (2) map read on the end of the adaptor labeled amplicons (*e.g.*, 96 genes and spike-in controls) that is not attached to the adaptor and/or barcode (*e.g.*, molecular barcode, sample tag, molecular identifier label); and (3) count and/or summarize the number of different adaptor labeled amplicons.

IDT Exhibit 1018.0886

[00118] FIG. 67 shows a schematic of a workflow for stochastically labeling nucleic acids with molecular barcodes (1220). As shown in step 1 of FIG. 67, RNA molecules may be stochastically labeled with a set of molecular barcodes (1220). The molecular barcodes (1220) may comprise a target binding region (1221), label region (1222), sample index region (1223) and universal PCR region (1224). In some instances, the target binding region comprises an oligodT sequence that hybridizes to a polyA sequence in the RNA molecules. The label region (1222) may contain a unique sequence that may be used to distinguish two or more different molecular barcodes. When the molecular barcode hybridizes to an RNA molecule, the label region may be used to confer a unique identity to identical RNA molecules. The sample index region (1223) may be identical for a set of molecular barcodes. The sample index region (1223) may be used to distinguish labeled nucleic acids from different samples. The universal PCR region (1224) may serve as a primer binding site for amplification of the labeled molecules. Once the RNA molecules are labeled with the molecular barcodes, the RNA molecules may be reverse transcribed to produce labeled cDNA molecules (1230) containing a cDNA copy of the RNA molecule (1210) and the molecular barcode (1220).

[00119] As shown in Step 2 of FIG. 67, excess oligos (*e.g.*, molecular barcodes) may be removed by Ampure bead purification. As shown in Step 3 of FIG. 67, the labeled cDNA molecules may be amplified by multiplex PCR. Multiplex PCR of the labeled cDNA molecules may be performed by using a first set of forward primers (F1, 1235 in FIG. 67) and universal primers (1240) in a single reaction volume to produce labeled amplicons (1245). As shown in Step 4 of FIG. 67, the labeled amplicons may be further amplified by multiplex PCR using nested primers. Nested primer amplification of the labeled amplicons may be performed by using a second set of forward primers (F2, 1250 in FIG. 67) and universal primers (1240) in a single reaction volume to produce labeled nested PCR amplicons. In some instances, the F2 primers (1250) contain an adaptor (1251) and a target binding region (1252). The target binding region (1252) of the F2 primers may hybridize to the labeled amplicons and may prime amplification of the labeled amplicons. The adaptor (1251) and the universal PCR region (1224) of the nested PCR amplicons may be used in the sequencing of the labeled nested PCR amplicons. The amplicons may be sequenced by MiSeq. Alternatively, the amplicons may be sequenced by HiSeq.

[00120] FIG. 68 shows a schematic of a workflow for stochastically labeling nucleic acids. As shown in Step 1 of FIG. 68, RNA molecules (1305) may be stochastically labeled with a set molecular barcodes (1320). The molecular barcodes may comprise a target binding region

IDT Exhibit 1018.0887

(1321), label region (1322), and universal PCR region (1323). Once the molecular barcodes are attached to the RNA molecules, the RNA molecules (1305) may be reverse transcribed to produce labeled cDNA molecules (1325) comprising a cDNA copy of the RNA molecule (1310) and the molecular barcode (1320). As shown in Step 2 of FIG. 68, the labeled cDNA molecules may be purified by Ampure bead purification to remove excess oligos (*e.g.*, molecular barcodes). As shown in Step 3 of FIG. 68, the labeled amplicons may be amplified by multiplex PCR. Multiplex PCR of the labeled cDNA molecules may be performed by using a first set of forward primers (F1, 1330 in FIG. 68) and universal primers (1335) in a single reaction volume to produce labeled amplicons (1360). As shown in Step 4 of FIG. 67, the labeled amplicons may be further amplified by multiplex PCR using nested primers. Nested primer amplification of the labeled amplicons may be performed by using a second set of forward primers (F2, 1340 in FIG. 68) and sample index primers (1350) in a single reaction volume to produce labeled nested PCR amplicons. In some instances, the F2 primers (1340) contain an adaptor (1341) and a target binding region (1342). The target binding region (1342) of the F2 primers may hybridize to the labeled amplicons and may prime amplification of the labeled amplicons. The sample index primers (1350) may comprise a universal primer region (1351), sample index region (1352), and adaptor region (1353). As shown in Step 4 of FIG. 68, the universal primer region (1351) of the sample index primer may hybridize to the universal PCR region of the labeled amplicons. The sample index region (1352) of the sample index primer may be used to distinguish two or more samples. The adaptor regions (1341, 1353) may be used to sequence the labeled nested PCR amplicons. The amplicons may be sequenced by MiSeq. Alternatively, the amplicons may be sequenced by HiSeq.

[00121] Further disclosed herein are methods of producing one or more libraries. The one or more libraries may comprise a plurality of labeled molecules. The one or more libraries may comprise a plurality of labeled amplicons. The one or more libraries may comprise a plurality of enriched molecules or a derivative thereof (*e.g.*, labeled molecules, labeled amplicons). Generally, the method of producing one or more libraries comprises (a) stochastically labeling a plurality of molecules from two or more samples to produce a plurality of labeled molecules, wherein the labeled molecules comprise a molecule region, a sample index region, and label region; and (b) producing one or more libraries from the plurality of labeled molecules, wherein (i) the one or more libraries comprise two or more different labeled molecules, (ii) the two or more different labeled molecules differ by the molecule region, sample index region, label region, or a combination thereof.

IDT Exhibit 1018.0888

[00122] The method for producing one or more libraries may comprise: a) producing a plurality of sample-tagged nucleic acids by: i) contacting a first sample comprising a plurality of nucleic acids with a plurality of first sample tags to produce a plurality of first sample-tagged nucleic acids; and ii) contacting a second sample comprising a plurality of nucleic acids with a plurality of second sample tags to produce a plurality of second sample-tagged nucleic acids, wherein the plurality of first sample tags are different from the second sample tags; and b) contacting the plurality of sample-tagged nucleic acids with a plurality of molecular identifier labels to produce a plurality of labeled nucleic acids, thereby producing a labeled nucleic acid library.

[00123] The contacting to a sample can be random or non-random. For example, the contacting of a sample with sample tags can be a random or non-random contacting. In some embodiments, the sample is contacted with sample tags randomly. In some embodiments, the sample is contacted with sample tags non-randomly. The contacting to a plurality of nucleic acids can be random or non-random. For example, the contacting of a plurality of nucleic acids with sample tags can be a random or non-random contacting. In some embodiments, the plurality of nucleic acids is contacted with sample tags randomly. In some embodiments, the plurality of nucleic acids is contacted with sample tags non-randomly.

[00124] Further disclosed herein are methods of producing one or more sets of labeled beads. The method of producing the one or more sets of labeled beads may comprise attaching one or more nucleic acids to one or more beads, thereby producing one or more sets of labeled beads. The one or more nucleic acids may comprise one or more molecular barcodes. The one or more nucleic acids may comprise one or more sample tags. The one or more nucleic acids may comprise one or more molecular identifier labels. The one or more nucleic acids may comprise a) a primer region; b) a sample index region; and c) a linker or adaptor region. The one or more nucleic acids may comprise a) a primer region; b) a label region; and c) a linker or adaptor region. The one or more nucleic acids may comprise a) a sample index region; and b) a label region. The one or more nucleic acids may further comprise a primer region. The one or more nucleic acids may further comprise a target specific region. The one or more nucleic acids may further comprise a linker region. The one or more nucleic acids may further comprise an adaptor region. The one or more nucleic acids may further comprise a sample index region. The one or more nucleic acids may further comprise a label region.

[00125] Further disclosed herein are methods for selecting one or more custom primers. The method of selecting a custom primer for analyzing molecules in a plurality of samples may

IDT Exhibit 1018.0889

comprise: a) a first pass, wherein primers chosen may comprise: i) no more than three sequential guanines, no more than three sequential cytosines, no more than four sequential adenines, and no more than four sequential thymines; ii) at least 3, 4, 5, or 6 nucleotides that are guanines or cytosines; and iii) a sequence that does not easily form a hairpin structure; b) a second pass, comprising: i) a first round of choosing a plurality of sequences that have high coverage of all transcripts; and ii) one or more subsequent rounds, selecting a sequence that has the highest coverage of remaining transcripts and a complementary score with other chosen sequences no more than 4; and c) adding sequences to a picked set until coverage saturates or total number of customer primers is less than or equal to about 96.

[00126] Further disclosed herein are kits for use in analyzing two or more molecules from two or more samples. The kit may comprise (a) a first container comprising a first set of molecular barcodes, wherein (i) a molecular barcode of the first set of molecular barcodes comprise a sample index region and a label region; (ii) the sample index region of two or more barcodes of the first set of molecular barcodes are the same; and (iii) the label region of two or more barcodes of the first set of molecular barcodes are different; and (b) a second container comprising a second set of molecular barcodes, wherein (i) a molecular barcode of the second set of molecular barcodes comprise a sample index region and a label region; (ii) the sample index region of two or more barcodes of the second set of molecular barcodes are the same; (iii) the label region of two or more barcodes of the second set of molecular barcodes are different; (iv) the sample index region of the barcodes of the second set of molecular barcodes are different from the sample index region of the barcodes of the first set of molecular barcodes; and (v) the label region of two or more barcodes of the second set of molecular barcodes are identical to the label region of two or more barcodes of the first set of molecular barcodes.

[00127] Alternatively, the kit comprises: a) a plurality of beads, wherein one or more beads of the plurality of beads may comprise at least one of a plurality of nucleic acids, wherein at least one of a plurality nucleic acids may comprise: i) at least one primer sequence, wherein the primer sequence of at least one of the plurality of nucleic acids is the same for the plurality of beads; ii) a bead-specific sequence, wherein the bead-specific sequence of any one of the plurality of nucleic acids is the same, and wherein the bead-specific sequence is different for any one of the plurality of beads; and iii) a stochastic sequence, wherein the stochastic sequence is different for any one of the plurality of nucleic acids; b) a primer may comprise a sequence complementary to the primer sequence; and c) one or more amplification agents suitable for nucleic acid amplification.

IDT Exhibit 1018.0890

[00128] Alternatively, the kit comprises: a) a first container comprising a first set of sample tags, wherein(i) a sample tag of the first set of sample tags comprises a sample index region; and (ii) the sample index regions of the sample tags of the first set of sample tags are at least about 80% identical; and b) a second container comprising a first set of molecular identifier labels, wherein (i) a molecular identifier label of the first set of molecular identifier labels comprises a label region; and (ii) at least about 30% of the label regions of the total molecular identifier labels of the first set of molecular identifier labels are different

[00129] Before the present methods, kits and compositions are described in greater detail, it is to be understood that this invention is not limited to particular method, kit or composition described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims. Examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.,* amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

[00130] Methods, kits and compositions are provided for stochastic labeling of nucleic acids in a plurality of samples or in a complex nucleic acid preparation. These methods, kits and compositions find use in unraveling mechanisms of cellular response, differentiation or signal transduction and in performing a wide variety of clinical measurements. These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the methods, kits and compositions as more fully described below.

[00131] The methods disclosed herein comprise attaching one or more molecular barcodes, sample tags, and/or molecular identifier labels to two or more molecules from two or more samples. The molecular barcodes, sample tags and/or molecular identifier labels may comprise one or more oligonucleotides. In some instances, attachment of molecular barcodes, sample tags, and/or molecular identifier labels to the molecules comprises stochastic labeling of the molecules. Methods for stochastically labeling molecules may be found, for example, in U.S. Serial Numbers 12/969,581 and 13/327,526. Generally, the stochastic labeling

IDT Exhibit 1018.0891

method comprises the random attachment of a plurality of the tag and label oligonucleotides to one or more molecules. The molecular barcodes, sample tags, and/or molecular identifier labels are provided in excess of the one or more molecules to be labeled. In stochastic labeling, each individual molecule to be labeled has an individual probability of attaching to the plurality of the molecular barcodes, sample tags, and/or molecular identifier labels. The probability of each individual molecule to be labeled attaching to a particular molecular barcodes, sample tags, and/or molecular identifier labels may be about the same as any other individual molecule to be labeled. Accordingly, in some instances, the probability of any of the molecules in a sample finding any of the tags and labels is assumed to be equal, an assumption that may be used in mathematical calculations to estimate the number of molecules in the sample. In some circumstances the probability of attaching may be manipulated by, for example electing tags and labels with different properties that would increase or decrease the binding efficiency of that molecular barcodes, sample tags, and/or molecular identifier labels with an individual molecule. The tags and labels may also be varied in numbers to alter the probability that a particular molecular barcodes, sample tags, and/or molecular identifier labels will find a binding partner during the stochastic labeling. For example, one label is overrepresented in a pool of labels, thereby increasing the chances that the overrepresented label finds at least one binding partner.

[00132] The methods disclosed herein may further comprise combining two or more samples. The methods disclosed herein may further comprise combining one or more molecules from two or more samples. For example, the methods disclosed herein comprise combining a first sample and a second sample. The two or more samples may be combined after conducting one or more stochastic labeling procedures. The two or more samples may be combined after attachment of one or more sets of molecular barcodes to two or more molecules from the two or more samples. The two or more samples may be combined after attachment of one or more sets of sample tags to two or more molecules from the two or more samples. The two or more samples may be combined after attachment of one or more sets of molecular identifier labels to two or more molecules from the two or more samples. For example, the first and second samples are combined prior to contact with the plurality of molecular identifier labels.

[00133] Alternatively, the two or more samples may be combined prior to conducting one or more stochastic labeling procedures. The two or more samples may be combined prior to attachment of one or more sets of molecular barcodes to two or more molecules from the two or more samples. The two or more samples may be combined prior to attachment of one or

more sets of sample tags to two or more molecules from the two or more samples. The two or more samples may be combined prior to attachment of one or more sets of molecular identifier labels to two or more molecules from the two or more samples.

[00134] The two or more samples may be combined after conducting one or more assays on two or more molecules or derivatives thereof (*e.g.*, labeled molecules, amplicons) from the two or more samples. The one or more assays may comprise one or more amplification reactions. The one or more assays may comprise one or more enrichment assays. The one or more assays may comprise one or more detection assays. For example, the first and second samples are combined after detecting the labeled nucleic acids.

[00135] The two or more samples may be combined prior to conducting one or more assays on two or more molecules or derivatives thereof (*e.g.*, labeled molecules, amplicons) from the two or more samples. The one or more assays may comprise one or more amplification reactions. The one or more assays may comprise one or more enrichment assays. The one or more assays may comprise one or more detection assays. For example, the first and second samples are combined prior to detecting the labeled nucleic acids.

Supports

[00136] The present disclosure comprises compositions and methods for multiplex sequence analysis from single cells. The methods and compositions of the present disclosure provide for the use of solid supports. In some instances, the methods, kits, and compositions disclosed herein comprise a support.

[00137] The terms "support", "solid support", "semi-solid support", and "substrate" may be used interchangeably and refer to a material or group of materials having a rigid or semi-rigid surface or surfaces. A support may refer to any surface that is transferable from solution to solution or forms a structure for conducting oligonucleotide-based assays. The support or substrate may be a solid support. Alternatively, the support is a non-solid support. A support may refer to an insoluble, semi-soluble, or insoluble material. A support may be referred to as "functionalized" when it includes a linker, a scaffold, a building block, or other reactive moiety attached thereto, whereas a solid support may be "nonfunctionalized" when it lack such a reactive moiety attached thereto. The support may be employed free in solution, such as in a microtiter well format; in a flow-through format, such as in a column; or in a dipstick.

[00138] The support or substrate may comprise a membrane, paper, plastic, coated surface, flat surface, glass, slide, chip, or any combination thereof. In many embodiments, at least one surface of the support may be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different compounds with, for example,

IDT Exhibit 1018.0893

wells, raised regions, pins, etched trenches, or the like. According to other embodiments, the solid support(s) may take the form of resins, gels, microspheres, or other geometric configurations. Alternatively, the solid support(s) comprises silica chips, microparticles, nanoparticles, plates, and arrays. Solid supports may include beads (*e.g.*, silica gel, controlled pore glass, magnetic beads, Dynabeads, Wang resin; Merrifield resin, Sephadex/Sepharose beads, cellulose beads, polystyrene beads etc.), capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold silver, aluminum, silicon and copper), glass supports, plastic supports, silicon supports, chips, filters, membranes, microwell plates, slides, or the like. plastic materials including multiwell plates or membranes (*e.g.*, formed of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), wafers, combs, pins or needles (*e.g.*, arrays of pins suitable for combinatorial synthesis or analysis) or beads in an array of pits or nanoliter wells of flat surfaces such as wafers (*e.g.,* silicon wafers), wafers with pits with or without filter bottoms.

[00139] Methods and techniques applicable to polymer (including protein) array synthesis have been described in U.S. Patent Pub. No. 20050074787, WO 00/58516, U.S. Patent Nos. 5,143,854, 5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,405,783, 5,424,186, 5,451,683, 5,482,867, 5,491,074, 5,527,681, 5,550,215, 5,571,639, 5,578,832, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,795,716, 5,831,070, 5,837,832, 5,856,101, 5,858,659, 5,936,324, 5,968,740, 5,974,164, 5,981,185, 5,981,956, 6,025,601, 6,033,860, 6,040,193, 6,090,555, 6,136,269, 6,269,846 and 6,428,752, in PCT Publication No. WO 99/36760 and WO 01/58593, which are all incorporated herein by reference in their entirety for all purposes. Patents that describe synthesis techniques in specific embodiments include U.S. Patent Nos. 5,412,087, 6,147,205, 6,262,216, 6,310,189, 5,889,165, and 5,959,098. Nucleic acid arrays are described in many of the above patents, but many of the same techniques may be applied to polypeptide arrays. Additional exemplary substrates are disclosed in U.S. Patent No. 5,744,305 and US Patent Pub. Nos. 20090149340 and 20080038559.

[00140] The attachment of the labeled nucleic acids to the support may comprise amine-thiol crosslinking, maleimide crosslinking, N-hydroxysuccinimide or N-hydroxysulfosuccinimide, Zenon or SiteClick. Attaching the labeled nucleic acids to the support may comprise attaching biotin to the plurality of labeled nucleic acids and coating the one or more beads with streptavadin.

[00141] In some instances, a solid support may comprise a molecular scaffold. Exemplary molecular scaffolds may include antibodies, antigens, affinity reagents, polypeptides, nucleic acids, cellular organelles, and the like. Molecular scaffolds may be linked together (*e.g.*, a

solid support may comprise a plurality of connected molecular scaffolds). Molecular scaffolds may be linked together by an amino acid linker, a nucleic acid linker, a small molecule linkage (*e.g.*, biotin and avidin), and/or a matrix linkage (*e.g.*, PEG or glycerol). Linkages may be non-covalent. Linkages may be covalent. In some instances, molecular scaffolds may not be linked. A plurality of individual molecular scaffolds may be used in the methods of the disclosure.

[00142] In some instances a support may comprise a nanoparticle. The nanoparticle may be a nickel, gold, silver, carbon, copper, silicate, platinum cobalt, zinc oxide, silicon dioxide crystalline, and/or silver nanoparticle. Alternatively, or additionally, the nanoparticle may be a gold nanoparticle embedded in a porous manganese oxide. The nanoparticle may be an iron nanoparticle. The nanoparticle may be a nanotetrapod studded with nanoparticles of carbon.

[00143] A support may comprise a polymer. A polymer may comprise a matrix. A matrix may further comprise one or more beads. A polymer may comprise PEG, glycerol, polysaccharide, or a combination thereof. A polymer may be a plastic, rubber, nylon, silicone, neoprene, and/or polystyrene. A polymer may be a natural polymer. Examples of natural polymers include, but are not limited to, shellac, amber, wool, silk, cellulose, and natural rubber. A polymer may be a synthetic polymer. Examples of synthetic polymers include, but are not limited to, synthetic rubber, phenol formaldehyde resin (or Bakelite), neoprene, nylon, polyvinyl chloride (PVC or vinyl), polystyrene, polyethylene, polypropylene, polyacrylonitrile, PVB, and silicone.

[00144] A support may be a semi-solid support. A support may comprise a gel (*e.g.*, a hydrogel). The terms "hydrogel", "gel" and the like, are used interchangeably herein and may refer to a material which is not a readily flowable liquid and not a solid but a gel which gel is comprised of from 0.5% or more and preferably less than 40% by weight of gel forming solute material and from 95% or less and preferably more than 55% water. The gels of the invention may be formed by the use of a solute which is preferably a synthetic solute (but could be a natural solute, *e.g.*, for forming gelatin) which forms interconnected cells which binds to, entrap, absorb and/or otherwise hold water and thereby create a gel in combination with water, where water includes bound and unbound water. The gel may be the basic structure of the hydrogel patch of the invention will include additional components beyond the gel forming solute material and water such as an enzyme and a salt which components are further described herein. The gel may be a polymer gel.

[00145] A solid support may comprise a structured nanostructure. For example, the structured nanostructure may comprise capture containers (*e.g.*, a miniaturized honeycomb)

which may comprise the oligonucleotides to capture the cell and/or contents of the cell. In some instances, structured nanostructures may not need the addition of exogenous reagents.

[00146] In some instances, the support comprises a bead. A bead may encompass any type of solid or hollow sphere, ball, bearing, cylinder, or other similar configuration composed of plastic, ceramic, metal, or polymeric material onto which a nucleic acid may be immobilized (*e.g.*, covalently or non-covalently). A bead may comprise nylon string or strings. A bead may be spherical in shape. A bead may be non-spherical in shape. Beads may be unpolished or, if polished, the polished bead may be roughened before treating, (*e.g.*, with an alkylating agent). A bead may comprise a discrete particle that may be spherical (*e.g.*, microspheres) or have an irregular shape. Beads may comprise a variety of materials including, but not limited to, paramagnetic materials, ceramic, plastic, glass, polystyrene, methylstyrene, acrylic polymers, titanium, latex, sepharose, cellulose, nylon and the like. A bead may be attached to or embedded into one or more supports. A bead may be attached to a gel or hydrogel. A bead may be embedded into a gel or hydrogel. A bead may be attached to a matrix. A bead may be embedded into a matrix. A bead may be attached to a polymer. A bead may be embedded into a polymer. The spatial position of a bead within the support (*e.g.*, gel, matrix, scaffold, or polymer) may be identified using the oligonucleotide present on the bead which serves as a location address. Examples of beads include, but are not limited to, streptavidin beads, agarose beads, magnetic beads, Dynabeads®, MACS® microbeads, antibody conjugated beads (*e.g.*, anti-immunoglobulin microbead), protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligodT conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and BcMag™ Carboxy-Terminated Magnetic Beads. The diameter of the beads may be about 5μm, 10μm, 20μm, 25μm, 30μm, 35μm, 40μm, 45μm or 50μm. A bead may refer to any three dimensional structure that may provide an increased surface area for immobilization of biological particles and macromolecules, such as DNA and RNA.

[00147] A support may be porous. A support may be permeable or semi-permeable. A support may be solid. A support may be semi-solid. A support may be malleable. A support may be flexible. In some instances, a support may be molded into a shape. For example, a support may be placed over an object and the support may take the shape of the object. In some instances, the support is placed over an organ and takes the shape of the organ. In some instances, the support is produced by 3D-printing.

[00148] The support (*e.g.*, beads, nanoparticles) may be at least about 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 100, 500, 1000, or 2000 or more micrometers in diameter. The solid

IDT Exhibit 1018.0896

supports (*e.g.*, beads) may be at most about 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 100, 500, 1000, or 2000 or more micrometers in diameter. The diameter of the bead may be about 20 microns.

[00149] In some instances, a solid support comprises a dendrimer. A dendrimer may be smaller than a bead. A dendrimer may be subcellular. A dendrimer may be less than 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, or 0.1 micron in diameter. A dendrimer may be less than 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, or 0.01 micron in diameterA dendrimer may comprise three major portions, a core, an inner shell, and an outer shell. A dendrimer may be synthesized to have different functionality in each of these portions. The different functionality of the portions of the dendrimer may control properties such as solubility, thermal stability, and attachment of compounds for particular applications. A dendrimer may be synthetically processed. A dendrimer may be synthesized by divergent synthesis. Divergent synthesis may comprise assembling a dendrimer from a multifunctional core, which is extended outward by a series of reactions. Divergent synthesis may comprise a series of Michael reactions. Alternatively, a dendrimer may be synthesized by convergent synthesis. Convergent synthesis may comprise building dendrimers from small molecules that end up at the surface of the sphere, and reactions may proceed inward and are eventually attached to a core. Dendrimers may also be prepared by click chemistry. Click chemistry may comprise Diels-Alder reactions, thiol-yne reactions, azide-alkyne reactions, or a combination thereof. Examples of dendrimers include, but are not limited to, poly(amidoamine) (PAMAM) dendrimer, PEG-core denderimer, phosphorous dendrimer, polypropylenimine dendrimer, and polylysine dendrimer. A dendrimer may be a chiral dendrimer. Alternatively, a dendrimer may be an achiral dendrimer.

[00150] A solid support may comprise a portion of a dendrimer. The portion of the dendrimer may comprise a dendron. A dendron may comprise monodisperse wedge-shaped dendrimer sections with multiple terminal groups and a single reaction function at the focal point. A solid support may comprise a polyester dendrom. Examples of dendrons include, but are not limited to, polyester-8-hydroxyl-1-acetylene bis-MPA dendron, polyester-16-hydroxyl-1-acetylene bis-MPA dendron, polyester-32-hydroxyl-1-acetylene bis-MPA dendron, polyester-8-hydroxyl-1-carboxyl bis-MPA dendron, polyester-16-hydroxyl-1-carboxyl bis-MPA dendron, and polyester-32-hydroxyl-1-carboxyl bis-MPA dendron.

[00151] A solid support may comprise a hyberbranched polymer. A hyperbranched polymer may comprise polydisperse dendritic macromolecules that possess dendrimer-like properties. Often, hyberbranched polymers are prepared in a single synthetic polymerization step. The

hyperbranched polymer may be based on 2,2-bis(hydroxymethyl)propanoic acid (bis-MPA) monomer. Examples of hyperbranched polymers include, but are not limited to, hyperbranched bis-MPA polyester-16-hydroxyl, hyperbranched bis-MPA polyester-32-hydroxyl, and hyperbranched bis-MPA polyester-64-hydroxyl.

[00152] The solid support may be an array or microarray. The solid support may comprise discrete regions. The solid support may be an addressable array. In some instances, the array comprises a plurality of probes fixed onto a solid surface. The plurality of probes enables hybridization of the labeled-molecule and/or labeled-amplicon to the solid surface. The plurality of probes comprises a sequence that is complementary to at least a portion of the labeled-molecule and/or labeled-amplicon. In some instances, the plurality of probes comprises a sequence that is complementary to at least a portion of the sample tag, molecular identifier label, nucleic acid, or a combination thereof. In other instances, the plurality of probes comprises a sequence that is complementary to the junction formed by the attachment of the sample tag or molecular identifier label to the nucleic acid.

[00153] The array may comprise one or more probes. The probes may be in a variety of formats. The array may comprise a probe comprising a sequence that is complementary to at least a portion of the target nucleic acid and a sequence that is complementary to the unique identifier region of a sample tag or molecular identifier label, wherein the sample tag or molecular identifier label comprises an oligonucleotide. The sequence that is complementary to at least a portion of the target nucleic acid may be attached to the array. The sequence that is complementary to the unique identifier region may be attached to the array. The array may comprise a first probe comprising a sequence that is complementary to at least a portion of the target nucleic acid and a second probe that is complementary to the unique identifier region. There are various ways in which a stochastically labeled nucleic acid may hybridize to the arrays. For example, the junction of the unique identifier region and the target nucleic acid of the stochastically labeled nucleic acid may hybridize to the probe on the array. There may be a gap in the regions of the stochastically labeled nucleic acid that may hybridize to the probe on the array. Different regions of the stochastically labeled nucleic acid may hybridize to two or more probes on the array. Thus, the array probes may be in many different formats. The array probes may comprise a sequence that is complementary to a unique identifier region, a sequence that is complementary to the target nucleic acid, or a combination thereof. Hybridization of the stochastically labeled nucleic acid to the array may occur by a variety of ways. For example, two or more nucleotides of the stochastically labeled nucleic acid may hybridize to one or more probes on the array. The two or more

IDT Exhibit 1018.0898

nucleotides of the stochastically labeled nucleic acid that hybridize to the probes may be consecutive nucleotides, non-consecutive nucleotides, or a combination thereof. The stochastically labeled nucleic acid that is hybridized to the probe may be detected by any method known in the art. For example, the stochastically labeled nucleic acids may be directly detected. Directly detecting the stochastically labeled nucleic acid may comprise detection of a fluorophore, hapten, or detectable label. The stochastically labeled molecules may be indirectly detected. Indirect detection of the stochastically labeled nucleic acid may comprise ligation or other enzymatic or non-enzymatic methods.

[00154] The array may be in a variety of formats. For example, the array may be in a 16-, 32-, 48-, 64-, 80-, 96-, 112-, 128-, 144-, 160-, 176-, 192-, 208-, 224-, 240-, 256-, 272-, 288-, 304-, 320-, 336-, 352-, 368-, 384-, or 400-format. Alternatively, the array is in an 8x60K, 4x180K, 2x400K, 1x1M format. In other instances, the array is in an 8x15K, 4x44K, 2x105K, 1x244K format.

[00155] The array may comprise a single array. The single array may be on a single substrate. Alternatively, the array is on multiple substrates. The array may comprise multiple formats. The array may comprise a plurality of arrays. The plurality of arrays may comprise two or more arrays. For example, the plurality of arrays may comprise at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 arrays. In some instances, at least two arrays of the plurality of arrays are identical. Alternatively, at least two arrays of the plurality of arrays are different.

[00156] In some instances, the array comprises symmetrical chambered areas. For example, the array comprises 0.5 x 0.5 millimeters (mm), 1 x 1 mm, 1.5 x 1.5 mm, 2 x 2 mm, 2.5 x 2.5 mm, 3 x 3 mm, 3.5 x 3.5 mm, 4 x 4 mm, 4.5 x 4.5 mm, 5 x 5 mm, 5.5 x 5.5 mm, 6 x 6 mm, 6.5 x 6.5 mm, 7 x 7 mm, 7.5 x 7.5 mm, 8 x 8 mm, 8.5 x 8.5 mm, 9 x 9 mm, 9.5 x 9.5 mm, 10 x 10 mm, 10.5 x 10.5 mm, 11 x 11 mm, 11.5 x 11.5 mm, 12 x 12 mm, 12.5 x 12.5 mm, 13 x 13 mm, 13.5 x 13.5 mm, 14 x 14 mm, 14.5 x 14.5 mm, 15 x 15 mm, 15.5 x 15.5 mm, 16 x 16 mm, 16.5 x 16.5 mm, 17 x 17 mm, 17.5 x 17.5 mm, 18 x 18 mm, 18.5 x 18.5 mm, 19 x 19 mm, 19.5 x 19.5 mm, or 20 x 20 mm chambered areas. In some instances, the array comprises 6.5 x 6.5 mm chambered areas. Alternatively, the array comprises asymmetrical chambered areas. For example, the array comprises 6.5 x 0.5 mm, 6.5 x 1 mm, 6.5 x 1.5 mm, 6.5 x 2 mm, 6.5 x 2.5 mm, 6.5 x 3 mm, 6.5 x 3.5 mm, 6.5 x 4 mm, 6.5 x 4.5 mm, 6.5 x 5 mm, 6.5 x 5.5 mm, 6.5 x 6 mm, 6.5 x 6.5 mm, 6.5 x 7 mm, 6.5 x 7.5 mm, 6.5 x 8 mm, 6.5 x 8.5 mm, 6.5 x 9 mm, 6.5 x 9.5 mm, 6.5 x 10 mm, 6.5 x 10.5 mm, 6.5 x 11 mm, 6.5 x 11.5 mm, 6.5 x 12 mm, 6.5 x 12.5 mm, 6.5 x 13 mm, 6.5 x 13.5 mm, 6.5 x 14 mm, 6.5 x 14.5 mm, 6.5 x

15 mm, 6.5 x 15.5 mm, 6.5 x 16 mm, 6.5 x 16.5 mm, 6.5 x 17 mm, 6.5 x 17.5 mm, 6.5 x 18 mm, 6.5 x 18.5 mm, 6.5 x 19 mm, 6.5 x 19.5 mm, or 6.5 x 20 mm chambered areas.

[00157] The array may comprise at least about 1 micron (μm), 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm, 9 μm, 10 μm, 15 μm, 20 μm, 25 μm, 30 μm, 35 μm, 40 μm, 45 μm, 50 μm, 55 μm, 60 μm, 65 μm, 70 μm, 75 μm, 80 μm, 85 μm, 90 μm, 95 μm, 100 μm, 125 μm, 150 μm, 175 μm, 200 μm , 225 μm, 250 μm, 275 μm, 300 μm, 325 μm, 350 μm, 375 μm, 400 μm , 425 μm, 450 μm, 475 μm, or 500 μm spots. In some instances, the array comprises 70 μm spots.

[00158] The array may comprise at least about 1 μm, 2 μm, 3 μm, 4 μm, 5 μm, 6 μm, 7 μm, 8 μm, 9 μm, 10 μm, 15 μm, 20 μm, 25 μm, 30 μm, 35 μm, 40 μm, 45 μm, 50 μm, 55 μm, 60 μm, 65 μm, 70 μm, 75 μm, 80 μm, 85 μm, 90 μm, 95 μm, 100 μm, 125 μm, 150 μm, 175 μm, 200 μm , 225 μm, 250 μm, 275 μm, 300 μm, 325 μm, 350 μm, 375 μm, 400 μm , 425 μm, 450 μm, 475 μm, 500 μm, 525 μm, 550 μm, 575 μm, 600 μm , 625 μm, 650 μm, 675 μm, 700 μm, 725 μm, 750 μm, 775 μm, 800 μm , 825 μm, 850 μm, 875 μm, 900 μm, 925 μm, 950 μm, 975 μm, 1000 μm feature pitch. In some instances, the array comprises 161 μm feature pitch.

[00159] The array may comprise one or more probes. In some instances, the array comprises at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 probes. Alternatively, the array comprises at least about 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, or 3000 probes. The array may comprise at least about 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, or 10000 probes. In some instances, the array comprises at least about 960 probes. Alternatively, the array comprises at least about 2780 probes. The probes may be specific for the plurality of oligonucleotide tags. The probes may be specific for at least a portion of the plurality of oligonucleotide tags. The probes may be specific for at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97% or 100% of the total number of the plurality of oligonucleotide tags. Alternatively, the probes are specific for at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97% or 100% of the total number of different oligonucleotide tags of the plurality of oligonucleotide tags. The probes may be oligonucleotides. The oligonucleotides may be at least about 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 nucleotides long. In other instances, the probes are non-specific probes. For example, the probes may be specific for a detectable label that is attached to the labeled-molecule. The probe may be streptavidin.

IDT Exhibit 1018.0900

[00160] The array may be a printed array. In some instances, the printed array comprises one or more oligonucleotides attached to a substrate. For example, the printed array comprises 5' amine modified oligonucleotides attached to an epoxy silane substrate.

[00161] Alternatively, the array comprises a slide with one or more wells. The slide may comprise at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 wells. Alternatively, the slide comprises at least about 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 650, 700, 750, 800, 850, 900, 950, or 1000 wells. In some instances, the slide comprises 16 wells. Alternatively, the slide comprises 96 wells. In other instances, the slide comprises at least about 80, 160, 240, 320, 400, 480, 560, 640, 720, 800, 880, or 960 wells.

[00162] In some instances, the solid support is an Affymetrix 3K tag array, Arrayjet non-contact printed array, or Applied Microarrays Inc (AMI) array. Alternatively, the support comprises a contact printer, impact printer, dot printer, or pin printer.

[00163] The solid support may comprise the use of beads that self-assemble in microwells. For example, the solid support comprises Illumina's BeadArray Technology. Alternatively, the solid support comprises Abbott Molecular's Bead Array technology, and Applied Microarray's FlexiPlexTM system.

[00164] In other instances, the solid support is a plate. Examples of plates include, but are not limited to, MSD multi-array plates, MSD Multi-Spot® plates, microplate, ProteOn microplate, AlphaPlate, DELFIA plate, IsoPlate, and LumaPlate.

[00165] The method may further comprise attaching at least one of a plurality of labeled nucleic acids to a support. The support may comprise a plurality of beads. The support may comprise an array. The support may comprise a glass slide.

[00166] The glass slide may comprise one or more wells. The one or more wells may be etched on the glass slide. The one or more wells may comprise at least 960 wells. The glass slide may comprise one or more probes. The one or more probes may be printed onto the glass slide. The one or more wells may further comprise one or more probes. The one or more probes may be printed within the one or more wells. The one or more probes may comprise 960 nucleic acids.

[00167] The methods and kits disclosed herein may further comprise distributing the plurality of first sample tags, the plurality of second sample tags, the plurality of molecular identifier labels, or any combination thereof in a microwell plate. The methods and kits disclosed herein may further comprise distributing one or more beads in the microwell plate. The methods and kits disclosed herein may further comprise distributing the plurality of

IDT Exhibit 1018.0901

samples in a plurality of wells of a microwell plate. The one or more of the plurality of samples may comprise a plurality of cells. One or more of the plurality of samples may comprise a plurality of nucleic acids. The method may further comprise distributing one or fewer cells to the plurality of wells. The plurality of cells may be lysed in the microwell plate. The method may further comprise synthesizing cDNA in the microwell plate. Synthesizing cDNA may comprise reverse transcription of mRNA. The microwell plate may comprise a microwell plate fabricated on PDMS by soft lithography, etched on a silicon wafer, etched on a glass slide, patterned photoresist on a glass slide, or a combination thereof. The microwell may comprise a hole on a microcapillary plate. The microwell plate may comprise a water-in-oil emulsion. The microwell plate may comprise at least one or more wells. The microwell plate may comprise at least about 6 wells, 12 wells, 48 wells, 96 wells, 384 wells, 960 wells or 1000 wells.

[00168] The methods and kits may further comprise a chip. The microwell plate may be attached to the chip. The chip may comprise at least about 6 wells, 12 wells, 48 wells, 96 wells, 384 wells, 960 wells, 1000 wells, 2000 wells, 3000 wells, 4000 wells, 5000 wells, 6000 wells, 7000 wells, 8000 wells, 9000 wells, 10,000 wells, 20,000 wells, 30,000 wells, 40,000 wells, 50,000 wells, 60,000 wells, 70,000 wells, 80, 000 wells, 90,000 wells, 100,000 wells, 200,000 wells, 500,000 wells, or a million wells. The wells may comprise an area of at least about 300 microns^2, 400 microns^2, 500 microns^2, 600 microns^2, 700 microns^2, 800 microns^2, 900 microns^2, 1000 microns^2, 1100 microns^2, 1200 microns^2, 1300 microns^2, 1400 microns^2, 1500 microns^2. The method may further comprise distributing between about 10,000 and 30,000 samples on the chip.

Functionalized surfaces and oligonucleotides

[00169] The bead may comprise a functionalized surface. A functionalized surface may refer to the surface of the solid support comprising a functional group. A functional group may be a group capable of forming an attachment with another functional group. For example, a functional group may be biotin, which may form an attachment with streptavidin, another functional group. Exemplary functional groups may include, but are not limited to, aldehydes, ketones, carboxy groups, amino groups, biotin, streptavidin, nucleic acids, small molecules (*e.g.*, for click chemistry), homo- and hetero-bifunctional reagents (*e.g.*, N-succinimidyl(4-iodoacetyl) aminobenzoate (SIAB), dimaleimide, dithio-bis-nitrobenzoic acid (DTNB), N-succinimidyl-S-acetyl-thioacetate (SATA), N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP), succinimidyl 4-(N-mafeimidomethyl)-cyclohexane-1-carboxylate (SMCC) and 6-

IDT Exhibit 1018.0902

hydrazinonicotimide (HYNIC), and antibodies. In some instances the functional group is a carboxy group (*e.g.*, COOH).

[00170] Oligonucleotides (*e.g.*, nucleic acids) may be attached to functionalized solid supports. The immobilized oligonucleotides on solid supports or similar structures may serve as nucleic acid probes, and hybridization assays may be conducted wherein specific target nucleic acids may be detected in complex biological samples.

[00171] The solid support (*e.g.*, beads) may be functionalized for the immobilization of oligonucleotides. An oligonucleotide may be conjugated to a solid support through a covalent amide bond formed between the solid support and the oligonucleotide.

[00172] A support may be conjugated to at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000 or more oligonucleotides. A support may be conjugated to at least about 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, 1000000, 2000000, 3000000, 4000000, 5000000, 6000000, 7000000, 8000000, 9000000 or 10000000, 100000000, 500000000, 1000000000 or more oligonucleotides. A support may be conjugated to at least about 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, 1000000, 2000000, 3000000, 4000000, 5000000, 6000000, 7000000, 8000000, 9000000 or 10000000, 100000000, 500000000, 1000000000 or more oligonucleotides. A support may be conjugated to at least 1 million oligonucleotides. A support may be conjugated to at least 10 million oligonucleotides. A support may be conjugated to at least 25 million oligonucleotides. A support may be conjugated to at least 50 million oligonucleotides. A support may be conjugated to at least 100 million oligonucleotides. A support may be conjugated to at least 250 million oligonucleotides. A support may be conjugated to at least 500 million oligonucleotides. A support may be conjugated to at least 750 million oligonucleotides. A support may be conjugated to at least about 1, 2, 3 4, 5, 6, 7, 8, 9, 10, 11, 12 13, 14, or 15 billion oligonucleotides. A support may be conjugated to at least 1 billion oligonucleotides. A support may be conjugated to at least 5 billion oligonucleotides.

[00173] The oligonucleotides may be attached to the support (*e.g.*, beads, polymers, gels) via a linker. Conjugation may comprise covalent or non-covalent attachment. Conjugation may introduce a variable spacer between the beads and the nucleic acids. The linker between the support and the oligonucleotide may be cleavable (*e.g.*, photocleavable linkage, acid labile linker, heat sensitive linker, and enzymatically cleavable linker).

[00174] Cross-linking agents for use for conjugating molecules to supports may include agents capable of reacting with a functional group present on a surface of the solid support

IDT Exhibit 1018.0903

and with a functional group present in the molecule. Reagents capable of such reactivity may include aldehydes, ketones, carboxy groups, amino groups, biotin, streptavidin, nucleic acids, small molecules (*e.g.*, for click chemistry), homo- and hetero-bifunctional reagents (*e.g.*, N-succinimidyl(4-iodoacetyl) aminobenzoate (SIAB), dimaleimide, dithio-bis-nitrobenzoic acid (DTNB), N-succinimidyl-S-acetyl-thioacetate (SATA), N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP), succinimidyl 4-(N-mafeimidomethyl)-cyclohexane-1-carboxylate (SMCC) and 6-hydrazinonicotimide (HYNIC).

[00175] A bead may be functionalized with a carboxy functional group and an oligonucleotide may be functionalized with an amino functional group.

[00176] A support may be smooth. Alternatively, or additionally, a support may comprise divets, ridges, or wells. A support may comprise a microwell array. A microwell array may be functionalized with functional groups that facilitate the attachment of oligonucleotides. The functional groups on the microwell array may be different for different positions on the microwell array. The functional groups on the microwell array may be the same for all regions of the microwell array.

Assay System Components

Microwell arrays

[00177] As described above, microwell arrays are used to entrap single cells and beads (one bead per cell) within a small reaction chamber of defined volume. Each bead comprises a library of oligonucleotide probes for use in stochastic labeling and digital counting of the entire complement of cellular mRNA molecules, which are released upon lysis of the cell. In one embodiment of the present disclosure, the microwell arrays are a consumable component of the assay system. In other embodiments, the microwell arrays may be reusable. In either case, they may be configured to be used as a stand-alone device for use in performing assays manually, or they may be configured to comprise a removable or fixed component of an instrument that provides for full or partial automation of the assay procedure.

[00178] The microwells of the array can be fabricated in a variety of shapes and sizes, which are chosen to optimize the efficiency of trapping a single cell and bead in each well. Appropriate well geometries include, but are not limited to, cylindrical, conical, hemispherical, rectangular, or polyhedral (*e.g.*, three dimensional geometries comprised of several planar faces, for example, hexagonal columns, octagonal columns, inverted triangular pyramids, inverted square pyramids, inverted pentagonal pyramids, inverted hexagonal pyramids, or inverted truncated pyramids). The microwells may comprise a shape that combines two or more of these geometries. For example, in one embodiment it may be partly

IDT Exhibit 1018.0904

cylindrical, with the remainder having the shape of an inverted cone. In another embodiment, it may include two side-by-side cylinders, one of larger diameter than the other, that are connected by a vertical channel (that is, parallel to the cylinder axes) that extends the full length (depth) of the cylinders. In general, the open end (or mouth) of each microwell will be located at an upper surface of the microwell array, but in some embodiments the openings may be located at a lower surface of the array. In general, the closed end (or bottom) of the microwell will be flat, but curved surfaces (*e.g.*, convex or concave) are also possible. In general, the shape (and size) of the microwells will be determined based on the types of cells and/or beads to be trapped in the microwells.

[00179] Microwell dimensions may be characterized in terms of the diameter and depth of the well. As used herein, the diameter of the microwell refers to the largest circle that can be inscribed within the planar cross-section of the microwell geometry. In one embodiment of the present disclosure, the diameter of the microwells may range from about 0.1 to about 5-fold the diameter of the cells and/or beads to be trapped within the microwells. In other embodiments, the microwell diameter is at least 0.1-fold, at least 0.5-fold, at least 1-fold, at least 2-fold, at least 3-fold, at least 4-fold, or at least 5-fold the diameter of the cells and/or beads to be trapped within the microwells. In yet other embodiments, the microwell diameter is at most 5-fold, at most 4-fold, at most 3-fold, at most 2-fold, at most 1-fold, at most 0.5-fold, or at most 0.1-fold the diameter of the cells and/or beads to be trapped within the microwells. In one embodiment, the microwell diameter is about 2.5-fold the diameter of the cells and/or beads to be trapped within the microwells. Those of skill in the art will appreciate that the microwell diameter may fall within any range bounded by any of these values (*e.g.,* from about 0.2-fold to about 3.5-fold the diameter of the cells and/or beads to be trapped within the microwells). Alternatively, the diameter of the microwells can be specified in terms of absolute dimensions. In one embodiment of the present disclosure, the diameter of the microwells may range from about 5 to about 50 microns. In other embodiments, the microwell diameter is at least 5 microns, at least 10 microns, at least 15 microns, at least 20 microns, at least 25 microns, at least 30 microns, at least 35 microns, at least 40 microns, at least 45 microns, or at least 50 microns. In yet other embodiments, the microwell diameter is at most 50 microns, at most 45 microns, at most 40 microns, at most 35 microns, at most 30 microns, at most 25 microns, at most 20 microns, at most 15 microns, at most 10 microns, or at most 5 microns. In one embodiment, the microwell diameter is about 30 microns. Those of skill in the art will appreciate that the microwell diameter may fall

within any range bounded by any of these values (*e.g.,* from about 28 microns to about 34 microns).

[00180] The microwell depth is chosen to optimize cell and bead trapping efficiency while also providing efficient exchange of assay buffers and other reagents contained within the wells. In one embodiment of the present disclosure, the depth of the microwells may range from about 0.1 to about 5-fold the diameter of the cells and/or beads to be trapped within the microwells. In other embodiments, the microwell depth is at least 0.1-fold, at least 0.5-fold, at least 1-fold, at least 2-fold, at least 3-fold, at least 4-fold, or at least 5-fold the diameter of the cells and/or beads to be trapped within the microwells. In yet other embodiments, the microwell depth is at most 5-fold, at most 4-fold, at most 3-fold, at most 2-fold, at most 1-fold, at most 0.5-fold, or at most 0.1-fold the diameter of the cells and/or beads to be trapped within the microwells. In one embodiment, the microwell depth is about 2.5-fold the diameter of the cells and/or beads to be trapped within the microwells. Those of skill in the art will appreciate that the microwell depth may fall within any range bounded by any of these values (*e.g.,* from about 0.2-fold to about 3.5-fold the diameter of the cells and/or beads to be trapped within the microwells). Alternatively, the diameter of the microwells can be specified in terms of absolute dimensions. In one embodiment of the present disclosure, the depth of the microwells may range from about 10 to about 60 microns. In other embodiments, the microwell depth is at least 10 microns, at least 20 microns, at least 25 microns, at least 30 microns, at least 35 microns, at least 40 microns, at least 50 microns, or at least 60 microns. In yet other embodiments, the microwell depth is at most 60 microns, at most 50 microns, at most 40 microns, at most 35 microns, at most 30 microns, at most 25 microns, at most 20 microns, or at most 10 microns. In one embodiment, the microwell depth is about 30 microns. Those of skill in the art will appreciate that the microwell depth may fall within any range bounded by any of these values (*e.g.,* from about 24 microns to about 36 microns).

[00181] The wells of the microwell array are arranged in a one dimensional, two dimensional, or three dimensional array, where three dimensional arrays may be achieved, for example, by stacking a series of two or more two dimensional arrays (that is, by stacking two or more substrates comprising microwell arrays). The pattern and spacing between wells is chosen to optimize the efficiency of trapping a single cell and bead in each well, as well as to maximize the number of wells per unit area of the array. The wells may be distributed according to a variety of random or non-random patterns, for example, they may be distributed entirely randomly across the surface of the array substrate, or they may be

IDT Exhibit 1018.0906

arranged in a square grid, rectangular grid, or hexagonal grid. In one embodiment of the present disclosure, the center-to-center distance (or spacing) between wells may vary from about 15 microns to about 75 microns. In other embodiments, the spacing between wells is at least 15 microns, at least 20 microns, at least 25 microns, at least 30 microns, at least 35 microns, at least 40 microns, at least 45 microns, at least 50 microns, at least 55 microns, at least 60 microns, at least 65 microns, at least 70 microns, or at least 75 microns. In yet other embodiments, the microwell spacing is at most 75 microns, at most 70 microns, at most 65 microns, at most 60 microns, at most 55 microns, at most 50 microns, at most 45 microns, at most 40 microns, at most 35 microns, at most 30 microns, at most 25 microns, at most 20 microns, or at most 15 microns. In one embodiment, the microwell spacing is about 55 microns. Those of skill in the art will appreciate that the microwell depth may fall within any range bounded by any of these values (*e.g.,* from about 18 microns to about 72 microns).

[00182] The microwell array may comprise surface features between the microwells that are designed to help guide cells and beads into the wells and/or prevent them from settling on the surfaces between wells. Examples of suitable surface features include, but are not limited to, domed, ridged, or peaked surface features that encircle the wells and/or straddle the surface between wells.

[00183] The total number of wells in the microwell array is determined by the pattern and spacing of the wells and the overall dimensions of the array. In one embodiment of the present disclosure, the number of microwells in the array may range from about 96 to about 5,000,000 or more. In other embodiments, the number of microwells in the array is at least 96, at least 384, at least 1,536, at least 5,000, at least 10,000, at least 25,000, at least 50,000, at least 75,000, at least 100,000, at least 500,000, at least 1,000,000, or at least 5,000,000. In yet other embodiments, the number of microwells in the array is at most 5,000,000, at most 1,000,000, at most 75,000, at most 50,000, at most 25,000, at most 10,000, at most 5,000, at most 1,536, at most 384, or at most 96 wells. In one embodiment, the number of microwells in the array is about 96. In another embodiment, the number of microwells is about 150,000. Those of skill in the art will appreciate that the number of microwells in the array may fall within any range bounded by any of these values (*e.g.,* from about 100 to 325,000).

[00184] Microwell arrays may be fabricated using any of a number of fabrication techniques known to those of skill in the art. Examples of fabrication methods that may be used include, but are not limited to, bulk micromachining techniques such as photolithography and wet chemical etching, plasma etching, or deep reactive ion etching;

IDT Exhibit 1018.0907

micro-molding and micro-embossing; laser micromachining; 3D printing or other direct write fabrication processes using curable materials; and similar techniques.

[00185] Microwell arrays may be fabricated from any of a number of substrate materials known to those of skill in the art, where the choice of material typically depends on the choice of fabrication technique, and vice versa. Examples of suitable materials include, but are not limited to, silicon, fused-silica, glass, polymers (*e.g.,* agarose, gelatin, hydrogels, polydimethylsiloxane (PDMS; elastomer), polymethylmethacrylate (PMMA), polycarbonate (PC), polypropylene (PP), polyethylene (PE), high density polyethylene (HDPE), polyimide, cyclic olefin polymers (COP), cyclic olefin copolymers (COC), polyethylene terephthalate (PET), and epoxy resins), metals or metal films (*e.g.,* aluminum, stainless steel, copper, nickel, chromium, and titanium), and the like. Typically, a hydrophilic material is desirable for fabrication of the microwell arrays (to enhance wettability and minimize non-specific binding of cells and other biological material), but hydrophobic materials that can be treated or coated (*e.g.,* by oxygen plasma treatment, or grafting of a polyethylene oxide surface layer) can also be used. The use of porous, hydrophilic materials for the fabrication of the microwell array may be desirable in order to facilitate capillary wicking/venting of entrapped air bubbles in the device. In some embodiments, the microwell array is fabricated with an optical adhesive. In some embodiments, the microwell array is fabricated with a plasma or corona treated material. The use of plasma or corona treated materials can make the material hydrophillic. In some embodients, plasma or corona treated materials, such as a hydrophillic material, can be more stable than non-treated materials. In some embodiments, the microwell array is fabricated from a single material. In other embodiments, the microwell array may comprise two or more different materials that have been bonded together or mechanically joined.

[00186] A variety of surface treatments and surface modification techniques may be used to alter the properties of microwell array surfaces. Examples include, but are not limited to, oxygen plasma treatments to render hydrophobic material surfaces more hydrophilic, the use of wet or dry etching techniques to smooth (or roughen) glass and silicon surfaces, adsorption and/or grafting of polyethylene oxide or other polymer layers to substrate surfaces to render them more hydrophilic and less prone to non-specific adsorption of biomolecules and cells, the use of silane reactions to graft chemically-reactive functional groups to otherwise inert silicon and glass surfaces, etc. Photodeprotection techniques can be used to selectively activate chemically-reactive functional groups at specific locations in the array structure, for example, the selective addition or activation of chemically-reactive functional groups such as

primary amines or carboxyl groups on the inner walls of the microwells may be used to covalently couple oligonucleotide probes, peptides, proteins, or other biomolecules to the walls of the microwells. In general, the choice of surface treatment or surface modification utilized will depend both on the type of surface property that is desired and on the type of material from which the microwell array is made.

[00187] In some embodiments, it may be advantageous to seal the openings of microwells during, for example, cell lysis steps, to prevent cross hybridization of target nucleic acid between adjacent microwells. A microwell may be sealed using a cap such as a solid support or a bead, where the diameter of the bead is larger than the diameter of the microwell. For example, a bead used as a cap can be at least about 10, 20, 30, 40, 50, 60, 70, 80 or 90% larger than the diameter of the microwell. Alternatively, a cap may be at most about 10, 20, 30, 40, 50, 60, 70, 80 or 90% larger than the diameter of the microwell.

[00188] A bead used as a cap may comprise cross-linked dextran beads (*e.g.*, Sephadex). Cross-linked dextran can range from about 10 micrometers to about 80 micrometers. The cross-linked dextran of the bead cap can be from 20 micrometers to about 50 micrometers. A cap can comprise, for example, inorganic nanopore membranes (*e.g.*, aluminum oxides), dialysis membranes, glass slides, coverslips,and/or hydrophilic plastic film (*e.g.*, film coated with a thin film of agarose hydrated with lysis buffer).

[00189] In some embodiments, the cap may allow buffer to pass into and out of the microwell, while preventing macromolecules (*e.g.*, nucleic acids) from migrating out of the well. A macromolecule of at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides can be blocked from migrating into or out of the microwell by the cap. A macromolecule of at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides can be blocked from migrating into or out of the microwell by the cap.

[00190] In some embodiments, a sealed microwell array can comprise a single layer of beads on top of the microwells. In some embodiments, a sealed microwell array can comprise multiple layers of beads on top of the microwells. A sealed microwell array can comprise about 1, 2, 3, 4, 5, or 6 or more layers of beads.

Mechanical fixtures

[00191] When performing multiplexed, single cell stochastic labeling/molecular indexing assays manually, it is convenient to mount the microwell array in a mechanical fixture to create a reaction chamber and facilitate the pipetting or dispensing of cell suspensions and assay reagents onto the array (**FIGs. 69** and **70**). In the example illustrated in **FIG. 69**, the

IDT Exhibit 1018.0909

fixture accepts a microwell array fabricated on a 1 mm thick substrate, and provides mechanical support in the form of a silicone gasket to confine the assay reagents to a reaction chamber that is 16 mm wide x 35 mm long x approximately 4 mm deep, thereby enabling the use of 800 microliters to 1 milliliter of cell suspension and bead suspension (comprising bead-based oligonucleotide labels) to perform the assay.

[00192] The fixture consists of rigid, machined top and bottom plates (*e.g.,* aluminum) and a compressible (*e.g.,* silicone, polydimethylsiloxane) gasket for creating the walls of the chamber or well. Design features include: (i) Chamfered aperture edges and clearance for rotating microscope objectives in and out of position as needed (for viewing the microwell array at different magnifications). (ii) Controlled compression of the silicone gasket to ensure uniform, repeatable formation of a leak-proof seal with the microwell array substrate. (iii) Captive fasteners for convenient operation. (iv) A locating clamp mechanism for secure and repeatable positioning of the array. (v) Convenient disassembly for removal of the array during rinse steps.

[00193] The top and bottom plates may be fabricated using any of a variety of techniques (*e.g.,* conventional machining, CNC machining, injection molding, 3D printing, etc.) using a variety of materials (*e.g.,* aluminum, anodized aluminum, stainless steel, teflon, polymethylmethacrylate (PMMA), polycarbonate (PC), or similar rigid polymer materials).

[00194] The silicone (polydimethylsiloxane; PDMS) gasket may be configured to create multiple chambers (see **FIG. 71**) in order to run controls and experiments (or replicate experiments, or multiple independent experiments) in parallel. The gasket is molded from PDMS or similar elastomeric material using a Teflon mold that includes draft angles for the vertical gasket walls to provide for good release characteristics. Alternatively, molds can be machined from aluminum or other materials (*e.g.*, black delrin, polyetherimide (ultem), etc.), and coated with Teflon if necessary to provide for good release characteristics. The gasket mold designs are inverted, i.e. so that the top surface of the molded part (i.e. the surface at the interface with a glass slide or silicon wafer used to cover the mold during casting) becomes the surface for creating a seal with the microwell array substrate during use, thereby avoiding potential problems with mold surface roughness and surface contamination in creating a smooth gasket surface (to ensure a leak-proof seal with the array substrate), and also providing for a flexible choice of substrate materials and the option of pre-assembly by using the microwell array substrate as a base during casting. The gasket mold designs may also include force focusing ridges at the boundaries of the well areas, i.e. the central mesa(s) in the mold (which form the well(s)) have raised ridges at the locations which become the perimeter

IDT Exhibit 1018.0910

of the well(s), so that a cover placed on top of the mold after filling rests on a small contact area at the precise location where good edge profile is critical for forming a leak-proof seal between the gasket and substrate during use.

Instrument Systems

[00195] The present disclosure also includes instrument systems and consumables to support the automation of multiplexed, single cell stochastic labeling/molecular indexing assays. Such systems may include consumable cartridges that incorporate microwell arrays integrated with flow cells, as well as the instrumentation necessary to provide control and analysis functionality such as (i) fluidics control, (ii) temperature control, (iii) cell and/or bead distribution and collection mechanisms, (iv) cell lysis mechanisms, (v) imaging capability, and (vi) image processing. In some embodiments, the input for the system comprises a cell sample and the output comprises a bead suspension comprising beads having attached oligonucleotides that incorporate sample tags, cell tags, and molecular indexing tags. In other embodiments, the system may include additional functionality, such as thermal cycling capability for performing PCR amplification, in which case the input for the system comprises a cell sample and the output comprises an oligonucleotide library resulting from amplification of the oligonucleotides incorporating sample tags, cell tags, and molecular indexing tags that were originally attached to beads. In yet other embodiments, the system may also include sequencing capability, with or without the need for oligonucleotide amplification, in which case the input for the system is a cell sample and the output comprises a dataset further comprising the sequences of all sample tag, cell tag, and molecular indexing tags associated with the target sequences of interest.

Microwell array flow cells

[00196] In many embodiments of the automated assay system, the microwell array substrate will be packaged within a flow cell that provides for convenient interfacing with the rest of the fluid handling system and facilitates the exchange of fluids, *e.g.,* cell and bead suspensions, lysis buffers, rinse buffers, etc., that are delivered to the microwell array. Design features may include: (i) one or more inlet ports for introducing cell samples, bead suspensions, and/or other assay reagents, (ii) one or more microwell array chambers designed to provide for uniform filling and efficient fluid-exchange while minimizing back eddies or dead zones, and (iii) one or more outlet ports for delivery of fluids to a sample collection point and/or a waste reservoir. In some embodiments, the design of the flow cell may include a plurality of microarray chambers that interface with a plurality of microwell arrays such that one or more cell samples may be processed in parallel. In some embodiments, the design

IDT Exhibit 1018.0911

of the flow cell may further include features for creating uniform flow velocity profiles, i.e. "plug flow", across the width of the array chamber to provide for more uniform delivery of cells and beads to the microwells, for example, by using a porous barrier located near the chamber inlet and upstream of the microwell array as a "flow diffuser", or by dividing each array chamber into several subsections that collectively cover the same total array area, but through which the divided inlet fluid stream flows in parallel. In some embodiments, the flow cell may enclose or incorporate more than one microwell array substrate. In some embodiments, the integrated microwell array / flow cell assembly may constitute a fixed component of the system. In some embodiments, the microwell array / flow cell assembly may be removable from the instrument.

[00197] In general, the dimensions of fluid channels and the array chamber(s) in flow cell designs will be optimized to (i) provide uniform delivery of cells and beads to the microwell array, and (ii) to minimize sample and reagent consumption. In some embodiments, the width of fluid channels will be between 50 microns and 20 mm. In other embodiments, the width of fluid channels may be at least 50 microns, at least 100 microns, at least 200 microns, at least 300 microns, at least 400 microns, at least 500 microns, at least 750 microns, at least 1 mm, at least 2.5 mm, at least 5 mm, at least 10 mm, or at least 20 mm. In yet other embodiments, the width of fluid channels may at most 20 mm, at most 10 mm, at most 5 mm, at most 2.5 mm, at most 1 mm, at most 750 microns, at most 500 microns, at most 400 microns, at most 300 microns, at most 200 microns, at most 100 microns, or at most 50 microns. In one embodiment, the width of fluid channels is about 2 mm. Those of skill in the art will appreciate that the width of the fluid channels may fall within any range bounded by any of these values (*e.g.,* from about 250 microns to about 3 mm).

[00198] In some embodiments, the depth of the fluid channels will be between 50 microns and 10 mm. In other embodiments, the depth of fluid channels may be at least 50 microns, at least 100 microns, at least 200 microns, at least 300 microns, at least 400 microns, at least 500 microns, at least 750 microns, at least 1 mm, at least 1.25 mm, at least 1.5 mm, at least 1.75 mm, at least 2 mm, at least 2.5 mm, at least 3 mm, at least 3.5 mm, at least 4 mm, at least 4.5 mm, at least 5 mm, at least 5.5 mm, at least 6 mm, at least 6.5 mm, at least 7 mm, at least 7.5 mm, at least 8 mm, at least 8.5 mm, at least 9 mm, or at least 9.5 mm. In other embodiments, the depth of fluid channels may be at most 10 mm, at most 9.5 mm, at most 9 mm, at most 8.5 mm, at most 8 mm, at most 7.5 mm, at most 7 mm, at most 6.5 mm, at most 6 mm, at most 5.5 mm, at most 5 mm, at most 4.5 mm, at most 4 mm, at most 3.5 mm, at most 3 mm, at most 2 mm, at most 1.75 mm, at most 1.5 mm, at most 1.25 mm, at most 1

mm, at most 750 microns, at most 500 microns, at most 400 microns, at most 300 microns, at most 200 microns, at most 100 microns, or at most 50 microns. In one embodiment, the depth of the fluid channels is about 1 mm. Those of skill in the art will appreciate that the depth of the fluid channels may fall within any range bounded by any of these values (*e.g.,* from about 800 microns to about 1 mm).

[00199] Flow cells may be fabricated using a variety of techniques and materials known to those of skill in the art. In general, the flow cell will be fabricated as a separate part and subsequently either mechanically clamped or permanently bonded to the microwell array substrate. Examples of suitable fabrication techniques include conventional machining, CNC machining, injection molding, 3D printing, alignment and lamination of one or more layers of laser or die-cut polymer films, or any of a number of microfabrication techniques such as photolithography and wet chemical etching, dry etching, deep reactive ion etching, or laser micromachining. Once the flow cell part has been fabricated it may be attached to the microwell array substrate mechanically, *e.g.,* by clamping it against the microwell array substrate (with or without the use of a gasket), or it may be bonded directly to the microwell array substrate using any of a variety of techniques (depending on the choice of materials used) known to those of skill in the art, for example, through the use of anodic bonding, thermal bonding, ultrasonic welding, or any of a variety of adhesives or adhesive films, including epoxy-based, acrylic-based, silicone-based, UV curable, polyurethane-based, or cyanoacrylate-based adhesives.

[00200] Flow cells may be fabricated using a variety of materials known to those of skill in the art. Examples of suitable materials include, but are not limited to, silicon, fused-silica, glass, any of a variety of polymers, *e.g.,* polydimethylsiloxane (PDMS; elastomer), polymethylmethacrylate (PMMA), polycarbonate (PC), polypropylene (PP), polyethylene (PE), high density polyethylene (HDPE), polyimide, cyclic olefin polymers (COP), cyclic olefin copolymers (COC), polyethylene terephthalate (PET), epoxy resins, metals (*e.g.,* aluminum, stainless steel, copper, nickel, chromium, and titanium), or a combination of these materials.

Cartridges

[00201] In many embodiments of the automated assay system, the microwell array, with or without an attached flow cell, will be packaged within a consumable cartridge that interfaces with the instrument system and which may incorporate additional functionality. Design features of cartridges may include (i) one or more inlet ports for creating fluid connections with the instrument and/or manually introducing cell samples, bead suspensions, and/or other

assay reagents into the cartridge, (ii) one or more bypass channels, i.e. for self-metering of cell samples and bead suspensions, to avoid overfilling and/or back flow, (iii) one or more integrated microwell array / flow cell assemblies, or one or more chambers within which the microarray substrate(s) are positioned, (iv) integrated miniature pumps or other fluid actuation mechanisms for controlling fluid flow through the device, (v) integrated miniature valves for compartmentalizing pre-loaded reagents and/or controlling fluid flow through the device, (vi) vents for providing an escape path for trapped air, (vii) one or more sample and reagent waste reservoirs, (viii) one or more outlet ports for creating fluid connections with the instrument and/or providing a processed sample collection point, (ix) mechanical interface features for reproducibly positioning the removable, consumable cartridge with respect to the instrument system, and for providing access so that external magnets can be brought into close proximity with the microwell array, (x) integrated temperature control components and/or a thermal interface for providing good thermal contact with the instrument system, and (xi) optical interface features, *e.g.,* a transparent window, for use in optical interrogation of the microwell array. In some embodiments, the cartridge is designed to process more than one sample in parallel. In some embodiments of the device, the cartridge may further comprise one or more removable sample collection chamber(s) that are suitable for interfacing with stand-alone PCR thermal cyclers and/or sequencing instruments. In some embodiments of the device, the cartridge itself is suitable for interfacing with stand-alone PCR thermal cyclers and/or sequencing instruments.

[00202] In some embodiments of the device, the cartridge may further comprise components that are designed to create physical and/or chemical barriers that prevent diffusion of (or increase path lengths and diffusion times for) large molecules in order to minimize cross-contamination between microwells. Examples of such barriers include, but are not limited to, a pattern of serpentine channels used for delivery of cells and beads to the microwell array, a retractable platen or deformable membrane that is pressed into contact with the surface of the microwell array substrate during lysis or incubation steps, the use of larger beads, *e.g.,* Sephadex beads as described previously, to block the openings of the microwells, or the release of an immiscible, hydrophobic fluid from a reservoir within the cartridge during lysis or incubation steps, to effectively separate and compartmentalize each microwell in the array. Any or all of these barriers, or an embodiment without such barriers, may be combined with raising the viscosity of the solution in and adjacent to the microwells, *e.g.,* through the addition of solution components such as glycerol or polyethylene glycol.

IDT Exhibit 1018.0914

[00203] In general, the dimensions of fluid channels and the array chamber(s) in cartridge designs will be optimized to (i) provide uniform delivery of cells and beads to the microwell array, and (ii) to minimize sample and reagent consumption. In some embodiments, the width of fluid channels will be between 50 microns and 20 mm. In other embodiments, the width of fluid channels may be at least 50 microns, at least 100 microns, at least 200 microns, at least 300 microns, at least 400 microns, at least 500 microns, at least 750 microns, at least 1 mm, at least 2.5 mm, at least 5 mm, at least 10 mm, or at least 20 mm. In yet other embodiments, the width of fluid channels may at most 20 mm, at most 10 mm, at most 5 mm, at most 2.5 mm, at most 1 mm, at most 750 microns, at most 500 microns, at most 400 microns, at most 300 microns, at most 200 microns, at most 100 microns, or at most 50 microns. In one embodiment, the width of fluid channels is about 2 mm. Those of skill in the art will appreciate that the width of the fluid channels may fall within any range bounded by any of these values (*e.g.,* from about 250 microns to about 3 mm).

[00204] In some embodiments, the depth of the fluid channels in cartridge designs will be between 50 microns and 10 mm. In other embodiments, the depth of fluid channels may be at least 50 microns, at least 100 microns, at least 200 microns, at least 300 microns, at least 400 microns, at least 500 microns, at least 750 microns, at least 1 mm, at least 1.25 mm, at least 1.5 mm, at least 1.75 mm, at least 2 mm, at least 2.5 mm, at least 3 mm, at least 3.5 mm, at least 4 mm, at least 4.5 mm, at least 5 mm, at least 5.5 mm, at least 6 mm, at least 6.5 mm, at least 7 mm, at least 7.5 mm, at least 8 mm, at least 8.5 mm, at least 9 mm, or at least 9.5 mm. In yet other embodiments, the depth of fluid channels may be at most 10 mm, at most 9.5 mm, at most 9 mm, at most 8.5 mm, at most 8 mm, at most 7.5 mm, at most 7 mm, at most 6.5 mm, at most 6 mm, at most 5.5 mm, at most 5 mm, at most 4.5 mm, at most 4 mm, at most 3.5 mm, at most 3 mm, at most 2 mm, at most 1.75 mm, at most 1.5 mm, at most 1.25 mm, at most 1 mm, at most 750 microns, at most 500 microns, at most 400 microns, at most 300 microns, at most 200 microns, at most 100 microns, or at most 50 microns. In one embodiment, the depth of the fluid channels is about 1 mm. Those of skill in the art will appreciate that the depth of the fluid channels may fall within any range bounded by any of these values (*e.g.,* from about 800 microns to about 1 mm).

[00205] Cartridges may be fabricated using a variety of techniques and materials known to those of skill in the art. In general, the cartridges will be fabricated as a series of separate component parts (**FIG. 72**) and subsequently assembled (**FIGs. 72** and **73**) using any of a number of mechanical assembly or bonding techniques. Examples of suitable fabrication techniques include, but are not limited to, conventional machining, CNC machining, injection

molding, thermoforming, and 3D printing. Once the cartridge components have been fabricated they may be mechanically assembled using screws, clips, and the like, or permanently bonded using any of a variety of techniques (depending on the choice of materials used), for example, through the use of thermal or ultrasonic bonding/welding or any of a variety of adhesives or adhesive films, including epoxy-based, acrylic-based, silicone-based, UV curable, polyurethane-based, or cyanoacrylate-based adhesives.

[00206] Cartridge components may be fabricated using any of a number of suitable materials, including but not limited to silicon, fused-silica, glass, any of a variety of polymers, *e.g.,* polydimethylsiloxane (PDMS; elastomer), polymethylmethacrylate (PMMA), polycarbonate (PC), polypropylene (PP), polyethylene (PE), high density polyethylene (HDPE), polyimide, cyclic olefin polymers (COP), cyclic olefin copolymers (COC), polyethylene terephthalate (PET), epoxy resins, or metals (*e.g.,* aluminum, stainless steel, copper, nickel, chromium, and titanium).

[00207] As described above, the inlet and outlet features of the cartridge may be designed to provide convenient and leak-proof fluid connections with the instrument, or may serve as open reservoirs for manual pipetting of samples and reagents into or out of the cartridge. Examples of convenient mechanical designs for the inlet and outlet port connectors include, but are not limited to, threaded connectors, swaged connectors, Luer lock connectors, Luer slip or "slip tip" connectors, press fit connectors, and the like. In some embodiments, the inlet and outlet ports of the cartridge may further comprise caps, spring-loaded covers or closures, phase change materials, or polymer membranes that may be opened or punctured when the cartridge is positioned in the instrument, and which serve to prevent contamination of internal cartridge surfaces during storage and/or which prevent fluids from spilling when the cartridge is removed from the instrument. As indicated above, in some embodiments the one or more outlet ports of the cartridge may further comprise a removable sample collection chamber that is suitable for interfacing with stand-alone PCR thermal cyclers and/or sequencing instruments.

[00208] As indicated above, in some embodiments the cartridge may include integrated miniature pumps or other fluid actuation mechanisms for control of fluid flow through the device. Examples of suitable miniature pumps or fluid actuation mechanisms include, but are not limited to, electromechanically- or pneumatically-actuated miniature syringe or plunger mechanisms, chemical propellants, membrane diaphragm pumps actuated pneumatically or by an external piston, pneumatically-actuated reagent pouches or bladders, or electro-osmotic pumps.

[00209] As described above, in some embodiments the cartridge may include miniature valves for compartmentalizing pre-loaded reagents and/or controlling fluid flow through the device. Examples of suitable miniature valves include, but are not limited to, one-shot "valves" fabricated using wax or polymer plugs that can be melted or dissolved, or polymer membranes that can be punctured; pinch valves constructed using a deformable membrane and pneumatic, hydraulic, magnetic, electromagnetic, or electromechanical (solenoid) acutation, one-way valves constructed using deformable membrane flaps, and miniature gate valves.

[00210] As indicated above, in some embodiments the cartridge may include vents for providing an escape path for trapped air. Vents may be constructed according to a variety of techniques known to those of skill in the art, for example, using a porous plug of polydimethylsiloxane (PDMS) or other hydrophobic material that allows for capillary wicking of air but blocks penetration by water. Vents may also be constructed as apertures through hydrophobic barrier materials, such that wetting to the aperture walls does not occur at the pressures used during operation.

[00211] In general, the mechanical interface features of the cartridge provide for easily removable but highly precise and repeatable positioning of the cartridge relative to the instrument system. Suitable mechanical interface features include, but are not limited to, alignment pins, alignment guides, mechanical stops, and the like. In some embodiments, the mechanical design features will include relief features for bringing external apparatus, *e.g.,* magnets or optical components, into close proximity with the microwell array chamber (**FIG. 72**).

[00212] In some embodiments, the cartridge will also include temperature control components or thermal interface features for mating to external temperature control modules. Examples of suitable temperature control elements include, but are not limited to, resistive heating elements, miniature infrared-emitting light sources, Peltier heating or cooling devices, heat sinks, thermistors, thermocouples, and the like. Thermal interface features will typically be fabricated from materials that are good thermal conductors (*e.g.,* copper, gold, silver, aluminium, etc.) and will typically comprise one or more flat surfaces capable of making good thermal contact with external heating blocks or cooling blocks.

[00213] In many embodiments, the cartridge will include optical interface features for use in optical imaging or spectroscopic interrogation of the microwell array. Typically, the cartridge will include an optically transparent window, *e.g.,* the microwell substrate itself or the side of the flow cell or microarray chamber that is opposite the microwell array,

fabricated from a material that meets the spectral requirements for the imaging or spectroscopic technique used to probe the microwell array. Examples of suitable optical window materials include, but are not limited to, glass, fused-silica, polymethylmethacrylate (PMMA), polycarbonate (PC), cyclic olefin polymers (COP), or cyclic olefin copolymers (COC). Typically, the cartridge will include a second optically transparent or translucent window or region which can be used to illuminate the microwell array in transverse, reflected, or oblique illumination orientations.

Instruments

[00214] The present disclosure also includes instruments for use in the automation of multiplexed, single cell stochastic labeling/molecular indexing assays. As indicated above, these instruments may provide control and analysis functionality such as (i) fluidics control, (ii) temperature control, (iii) cell and/or bead distribution and collection mechanisms, (iv) cell lysis mechanisms, (v) magnetic field control, (vi) imaging capability, and (vii) image processing. In some embodiments, the instrument system may comprise one or more modules (one possible embodiment of which is illustrated schematically in **FIG. 74**), where each module provides one or more specific functional feature sets to the system. In other embodiments, the instrument system may be packaged such that all system functionality resides within the same package. **FIG. 75** provides a schematic illustration of the process steps included in one embodiment of the automated system. As indicated above, in some embodiments, the system may comprise additional functional units, either as integrated components or as modular components of the system, that expand the functional capabilities of the system to include PCR amplification (or other types of oligonucleotide amplification techniques) and oligonucleotide sequencing.

[00215] In general, the instrument system will provide fluidics capability for delivering samples and/or reagents to the one or more microarray chamber(s) or flow cell(s) within one or more assay cartridge(s) connected to the system. Assay reagents and buffers may be stored in bottles, reagent and buffer cartridges, or other suitable containers that are connected to the cartridge inlets. The system may also include waste reservoirs in the form of bottles, waste cartridges, or other suitable waste containers for collecting fluids downstream of the assay cartridge(s). Control of fluid flow through the system will typically be performed through the use of pumps (or other fluid actuation mechanisms) and valves. Examples of suitable pumps include, but are not limited to, syringe pumps, programmable syringe pumps, peristaltic pumps, diaphragm pumps, and the like. In some embodiments, fluid flow through the system may be controlled by means of applying positive pneumatic pressure at the one or

IDT Exhibit 1018.0918

more inlets of the reagent and buffer containers, or at the inlets of the assay cartridge(s). In some embodiments, fluid flow through the system may be controlled by means of drawing a vacuum at the one or more outlets of the waste reservoirs, or at the outlets of the assay cartridge(s). Examples of suitable valves include, but are not limited to, check valves, electromechanical two-way or three-way valves, pneumatic two-way and three-way valves, and the like. In some embodiments, pulsatile flow may be applied during assay wash/rinse steps to facilitate complete and efficient exchange of fluids within the one or more microwell array flow cell(s) or chamber(s).

[00216] As indicated above, in some embodiments the instrument system may include mechanisms for further facilitating the uniform distribution of cells and beads over the microwell array. Examples of such mechanisms include, but are not limited to, rocking, shaking, swirling, recirculating flow, low frequency agitation (for example, using a rocker plate or through pulsing of a flexible (*e.g.,* silicone) membrane that forms a wall of the chamber or nearby fluid channel), or high frequency agitation (for example, through the use of piezoelectric transducers). In some embodiments, one or more of these mechanisms is utilized in combination with physical structures or features on the interior walls of the flow cell or array chamber, *e.g.,* mezzanine/top hat structures, chevrons, or ridge arrays, to facilitate mixing and/or to help prevent pooling of cells or beads within the array chamber. Flow-enhancing ribs on upper or lower surfaces of the flow cell or array chamber may be used to control flow velocity profiles and reduce shear across the microwell openings (i.e. to prevent cells or beads from being pulled out of the microwells during reagent exchange and rinse steps).

[00217] In some embodiments, the instrument system may include mechanical cell lysis capability as an alternative to the use of detergents or other reagents. Sonication using a high frequency piezoelectric transducer is one example of a suitable technique.

[00218] In some embodiments, the instrument system will include temperature control functionality for the purpose of facilitating the accuracy and reproducibility of assay results, for example, cooling of the microwell array flow cell or chamber may be advantageous for minimizing molecular diffusion between microwells. Examples of temperature control components that may be incorporated into the instrument system design include, but are not limited to, resistive heating elements, infrared light sources, Peltier heating or cooling devices, heat sinks, thermistors, thermocouples, and the like. In some embodiments of the system, the temperature controller may provide for programmable changes in temperature over specified time intervals.

IDT Exhibit 1018.0919

[00219] As indicated elsewhere in this disclosure, many embodiments of the disclosed methods utilize magnetic fields for removing beads from the microwells upon completion of the assay. In some embodiments, the instrument system may further comprise use of magnetic fields for transporting beads into or out of the microwell array flow cell or chamber. Examples of suitable means for providing control of magnetic fields include, but are not limited to, use of electromagnets in fixed position(s) relative to the cartridge, or the use of permanent magnets that are mechanically repositioned as necessary. In some embodiments of the instrument system, the strength of the applied magnetic field(s) will be varied by varying the amount of current applied to one or more electromagnets. In some embodiments of the instrument system, the strength of the applied magnetic fields will be varied by changing the position of one or more permanent magnets relative to the position of the microarray chamber(s) using, for example, stepper motor-driven linear actuators, servo motor-driven linear actuators, or cam shaft mechanisms. In some embodiments of the instrument system, the use of pulsed magnetic fields may be advantageous, for example, to prevent clustering of magnetic beads. In some embodiments, a magnet in close proximity to the array or chamber may be moved, once or multiple times, between at least two positions relative to the microwell array. Motion of the magnets can serve to agitate beads within microwells, to facilitate removal of beads from microwells, or to collect magnetic beads at a desired location.

[00220] As indicated above, in many embodiments the instrument system will include optical imaging and/or other spectroscopic capabilities. Such functionality may be useful, for example, for inspection of the microwell array(s) to determine whether or not the array has been uniformly and optimally populated with cells and/or beads. Any of a variety of imaging modes may be utilized, including but not limited to, bright-field, dark-field, and fluorescence/luminescence imaging. The choice of imaging mode will impact the design of microwell arrays, flow cells, and cartridge chambers in that the array substrate and/or opposing wall of the flow cell or array chamber will necessarily need to be transparent or translucent over the spectral range of interest. In some embodiments, each microwell array may be imaged in its entirety within a single image. In some embodiments, a series of images may be "tiled" to create a high resolution image of the entire array. In some embodiment, a single image that represents a subsection of the array may be used to evaluate properties, *e.g.,* cell or bead distributions, for the array as a whole. In some embodiments, dual wavelength excitation and emission (or multi-wavelength excitation and/or emission) imaging may be performed. Any of a variety of light sources may be used to provide the

imaging and/or excitation light, including but not limited to, tungsten lamps, tungsten-halogen lamps, arc lamps, lasers, light emitting diodes (LEDs), or laser diodes. Any of a variety of image sensors may be used for imaging purposes, including but not limited to, photodiode arrays, charge-coupled device (CCD) cameras, or CMOS image sensors. The optical system will typically include a variety of optical components for steering, shaping, filtering, and/or focusing light beams through the system. Examples of suitable optical components include, but are not limited to, lenses, mirrors, prisms, diffraction gratings, colored glass filters, narrowband interference filters, broadband interference filters, dichroic reflectors, optical fibers, optical waveguides, and the like. In some embodiments, the instrument system may use an optically transparent microarray substrate as a waveguide for delivering excitation light to the microwell array. The choice of imaging mode may also enable the use of other types of assays to be run in parallel with stochastic labeling / molecular indexing assays, for example, the use of trypan blue live cell / dead cell assays with bright field imaging, the use of fluorescence-based live cell / dead cell assays with fluorescence imaging, etc. Correlation of viability data for individual cells with the cell tag associated with each bead in the associated microwell may provide an additional level of discrimination in analyzing the data from multiplexed, single cell assays. Alternatively, viability data in the form of statistics for multiple cells may be employed for enhancing the analytical capabilities and quality assurance of the assay.

[00221] In some embodiments, the system may comprise non-imaging and/or non-optical capabilities for probing the microwell array. Examples of non-imaging and/or non-optical techniques for detecting trapped air bubbles, determining the cell and/or bead distribution over the array, etc., include but are not limited to measurements of light scattering, ultraviolet/visible/infrared absorption measurements (*e.g.*, using stained cells and/or beads that incorporate dyes), coherent raman scattering, and conductance measurements (*e.g.*, using microfabricated arrays of electrodes in register with the microwell arrays).

System processor and software

[00222] In general, instrument systems designed to support the automation of multiplexed, single cell stochastic labeling/molecular indexing assays will include a processor or computer, along with software to provide (i) instrument control functionality, (ii) image processing and analysis capability, and (iii) data storage, analysis, and display functionality.

[00223] In many embodiments, the instrument system will comprise a computer (or processor) and computer-readable media that includes code for providing a user interface as well as manual, semi-automated, or fully-automated control of all system functions, i.e.

IDT Exhibit 1018.0921

control of the fluidics system, the temperature control system, cell and/or bead distribution functions, magnetic bead manipulation functions, and the imaging system. Examples of fluid control functions provided by the instrument control software include, but are not limited to, volumetric fluid flow rates, fluid flow velocities, the timing and duration for sample and bead addition, reagent addition, and rinse steps. Examples of temperature control functions provided by the instrument control software include, but are not limited to, specifying temperature set point(s) and control of the timing, duration, and ramp rates for temperature changes. Examples of cell and/or bead distribution functions provided by the instrument control software include, but are not limited to, control of agitation parameters such as amplitude, frequency, and duration. Examples of magnetic field functions provided by the instrument control software include, but are not limited to, the timing and duration of the applied magnetic field(s), and in the case of electromagnets, the strength of the magnetic field as well. Examples of imaging system control functions provided by the instrument control software include, but are not limited to, autofocus capability, control of illumination and/or excitation light exposure times and intensities, control of image acquisition rate, exposure time, and data storage options.

[00224] In some embodiments of the instrument system, the system will further comprise computer-readable media that includes code for providing image processing and analysis capability. Examples of image processing and analysis capability provided by the software include, but are not limited to, manual, semi-automated, or fully-automated image exposure adjustment (*e.g.,* white balance, contrast adjustment, signal-averaging and other noise reduction capability, etc.), automated object identification (i.e. for identifying cells and beads in the image), automated statistical analysis (i.e. for determining the number of cells and/or beads identified per unit area of the microwell array, or for identifying wells that contain more than one cell or more than one bead), and manual measurement capabilities (*e.g.,* for measuring distances between objects, etc.). In some embodiments, the instrument control and image processing/analysis software will be written as separate software modules. In some embodiments, the instrument control and image processing/analysis software will be incorporated into an integrated package. In some embodiments, the system software may provide integrated real-time image analysis and instrument control, so that cell and bead sample loading steps can be prolonged or repeated until optimal cell/bead distributions are achieved.

[00225] In some embodiments of the instrument system, the system will comprise computer-readable media that includes code for providing sequence data analysis. Examples

IDT Exhibit 1018.0922

of sequence data analysis functionality that may be provided by the data analysis software includes, but is not limited to, (i) algorithms for determining the number of reads per gene per cell, and the number of unique transcript molecules per gene per cell, based on the data provided by sequencing the oligonucleotide library created by running the assay, (ii) statistical analysis of the sequencing data, *e.g.,* principal component analysis, for predicting confidence intervals for determinations of the number of transcript molecules per gene per cell, etc., (iii) sequence alignment capabilities for alignment of gene sequence data with known reference sequences, (iv) decoding/demultiplexing of sample barcodes, cell barcodes, and molecular barcodes, and (v) automated clustering of molecular labels to compensate for amplification or sequencing errors.

[00226] In general, the computer or processor included in the presently disclosed instrument systems, as illustrated in **FIG. 76**, may be further understood as a logical apparatus that can read instructions from media 511 and/or a network port 505, which can optionally be connected to server 509 having fixed media 512. The system 500, such as shown in **FIG. 76** can include a CPU 501, disk drives 503, optional input devices such as keyboard 515 and/or mouse 516 and optional monitor 507. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure can be transmitted over such networks or connections for reception and/or review by a party 522 as illustrated in **FIG. 76**.

[00227] **FIG. 77** is a block diagram illustrating a first example architecture of a computer system 100 that can be used in connection with example embodiments of the present disclosure. As depicted in **FIG. 77**, the example computer system can include a processor 102 for processing instructions. Non-limiting examples of processors include: Intel XeonTM processor, AMD OpteronTM processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0TM processor, ARM Cortex-A8 Samsung S5PC100TM processor, ARM Cortex-A8 Apple A4TM processor, Marvell PXA 930TM processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some embodiments, multiple processors or processors with multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[00228] As illustrated in **FIG. 77**, a high speed cache 104 can be connected to, or incorporated in, the processor 102 to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor 102. The processor 102 is connected to a north bridge 106 by a processor bus 108. The north bridge 106 is connected to random access memory (RAM) 110 by a memory bus 112 and manages access to the RAM 110 by the processor 102. The north bridge 106 is also connected to a south bridge 114 by a chipset bus 116. The south bridge 114 is, in turn, connected to a peripheral bus 118. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus 118. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip.

[00229] In some embodiments, system 100 can include an accelerator card 122 attached to the peripheral bus 118. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[00230] Software and data are stored in external storage 124 and can be loaded into RAM 110 and/or cache 104 for use by the processor. The system 100 includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, Windows™, MACOS™, BlackBerry OS™, iOS™, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example embodiments of the present invention.

[00231] In this example, system 100 also includes network interface cards (NICs) 120 and 121 connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[00232] **FIG. 78** is a diagram showing a network 200 with a plurality of computer systems 202a, and 202b, a plurality of cell phones and personal data assistants 202c, and Network Attached Storage (NAS) 204a, and 204b. In example embodiments, systems 212a, 212b, and 212c can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) 214a and 214b. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems 212a, and 212b, and

IDT Exhibit 1018.0924

cell phone and personal data assistant systems 212c. Computer systems 212a, and 212b, and cell phone and personal data assistant systems 212c can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) 214a and 214b. **FIG. 78** illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various embodiments of the present invention. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing. Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface.

[00233] In some example embodiments, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other embodiments, some or all of the processors can use a shared virtual address memory space.

[00234] **FIG. 79** is a block diagram of a multiprocessor computer system 300 using a shared virtual address memory space in accordance with an example embodiment. The system includes a plurality of processors 302a-f that can access a shared memory subsystem 304. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) 306a-f in the memory subsystem 304. Each MAP 306a-f can comprise a memory 308a-f and one or more field programmable gate arrays (FPGAs) 310a-f. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs 310a-f for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example embodiments. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory 308a-f, allowing it to execute tasks independently of, and asynchronously from, the respective microprocessor 302a-f. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[00235] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example embodiments, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some embodiments, all or part of the computer system can

IDT Exhibit 1018.0925

be implemented in software or hardware. Any variety of data storage media can be used in connection with example embodiments, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[00236] In example embodiments, the computer subsystem of the present disclosure can be implemented using software modules executing on any of the above or other computer architectures and systems. In other embodiments, the functions of the system can be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in **FIG. 79**, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer can be implemented with hardware acceleration through the use of a hardware accelerator card, such as accelerator card 122 illustrated in **FIG. 77**.

Oligonucleotides (*e.g.*, Molecular Barcodes)

[00237] The methods and kits disclosed herein may comprise one or more oligonucleotides or uses thereof. The oligonucleotides may be attached to a solid support disclosed herein. Attachment of the oligonucleotide to the solid support may occur through functional group pairs on the solid support and the oligonucleotide. The oligonucleotide may be referred to as a molecular bar code. The oligonucleotide may be referred to as a label (*e.g.*, molecular label, cellular label) or tag (*e.g.*, sample tag).

[00238] Oligonucleotides may comprise a universal label. A universal label may be the same for all oligonucleotides in a sample. A universal label may be the same for oligonucleotides in a set of oligonucleotides. A universal label may be the same for two or more sets of oligonucleotides. A universal label may comprise a sequence of nucleic acids that may hybridize to a sequencing primer. Sequencing primers may be used for sequencing oligonucleotides comprising a universal label. Sequencing primers (*e.g.*, universal sequencing primers) may comprise sequencing primers associated with high-throughput sequencing platforms. A universal label may comprise a sequence of nucleic acids that may hybridize to a PCR primer. A universal label may comprise a sequence of nucleic acids that may hybridize to a sequencing primer and a PCR primer. The sequence of nucleic acids of the universal label that may hybridize to a sequencing and/or PCR primer may be referred to as a primer binding site. A universal label may comprise a sequence that may be used to initiate transcription of the oligonucleotide. A universal label may comprise a sequence that may be used for extension of the oligonucleotide or a region within the oligonucleotide. A universal

label may be at least about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A universal label may comprise at least about 10 nucleotides. A universal label may be at most about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length.

[00239] Oligonucleotides may comprise a cellular label. A cellular label may comprise a nucleic acid sequence that may provide information for which cell the oligonucleotide is contacted to (*e.g.*, determining which nucleic acid originated from which cell). At least 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% or 100% of oligonucleotides on the same solid support may comprise the same cellular label. At least 60% of oligonucleotides on the same solid support may comprise the same cellular label. At least 95% of oligonucleotides on the same solid support may comprise the same cellular label. All the oligonucleotides on a same solid support may comprise the same cellular label. The cellular label of the oligonucleotides on a first solid support may be different than the cellular labels of the oligonucleotides on the second solid support.

[00240] A cellular label may be at least about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A cellular label may be at most about 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 or fewer or more nucleotides in length. A cellular label may comprise between about 5 to about 200 nucleotides. A cellular label may comprise between about 10 to about 150 nucleotides. A cellular label may comprise between about 20 to about 125 nucleotides in length.

[00241] Oligonucleotides may comprise a molecular label. A molecular label may comprise a nucleic acid sequence that may provide identifying information for the specific nucleic acid species hybridized to the oligonucleotide. Oligonucleotides conjugated to a same solid support may comprise different molecular labels. In this way, the molecular label may distinguish the types of target nucleic acids (*e.g.*, genes), that hybridize to the different oligonucleotides. A molecular label may be at least about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A molecular label may be at most about 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 or fewer nucleotides in length.

[00242] Oligonucleotides may comprise a sample label (*e.g.*, sample index). A sample label may comprise a nucleic acid sequence that may provide information about from where a target nucleic acid originated. For example, a sample label may be different on different solid supports used in different experiments. A sample label may be at least about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A sample label may be at most

IDT Exhibit 1018.0927

about 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 or fewer nucleotides in length.

[00243] An oligonucleotide may comprise a universal label, a cellular label, a molecular label and a sample label, or any combination thereof. In combination, the sample label may be used to distinguish target nucleic acids between samples, the cellular label may be used to distinguish target nucleic acids from different cells in the sample, the molecular label may be used to distinguish the different target nucleic acids in the cell (*e.g.*, different copies of the same target nucleic acid), and the universal label may be used to amplify and sequence the target nucleic acids.

[00244] A universal label, a molecular label, a cellular label, linker label and/or a sample label may comprise a random sequence of nucleotides. A random sequence of nucleotides may be computer generated. A random sequence of nucleotides may have no pattern associated with it. A universal label, a molecular label, a cellular label, linker label and/or a sample label may comprise a non-random (*e.g.*, the nucleotides comprise a pattern) sequence of nucleotides. Sequences of the universal label, a molecular label, a cellular label, linker label and/or a sample label may be commercially available sequences. Sequences of the universal label, a molecular label, a cellular label, linker label and/or a sample label may be comprise randomer sequences. Randomer sequences may refer to oligonucleotide sequences composed of all possible sequences for a given length of the randomer. Alternatively, or additionally, a universal label, a molecular label, a cellular label, linker label and/or a sample label may comprise a predetermined sequence of nucleotides.

[00245] **FIG. 1** shows an exemplary oligonucleotide of the disclosure comprising a universal label, a cellular label and a molecular label.

[00246] **FIG. 3** shows an exemplary oligonucleotide coupled solid support comprising a solid support (301) coupled to an oligonucleotide (312). The oligonucleotide (312) comprises a chemical group (5' amine, 302), a universal label (303), a cellular label (311), a molecular label (Molecular BC, 311), and a target binding region (oligodT, 310). In this schematic, the cellular label (311) comprises a first cell label (CL Part 1, 304), a first linker (Linker1, 305), a second cell label (CL Part 2, 306), a second linker (Linker2, 307), a third cell label (CL Part 3, 308). The cellular label (311) is common for each oligonucleotide on the solid support. The cellular labels (311) for two or more beads may be different. The cellular labels (311) for two or more beads may differ by the cell labels (*e.g.*, CL Part 1 (304), CL Part 2 (306), CL Part 3 (308)). The cellular labels (311) for two or more beads may differ by the first cell label (304), second cell label (306), third cell label (308), or a combination thereof. The first and second

IDT Exhibit 1018.0928

linkers (303, 305) of the cellular labels (311) may be identical for two or more oligonucleotide coupled solid supports. The universal label (303) may be identical for two or more oligonucleotide coupled solid supports. The universal label (303) may be identical for two or more oligonucleotides on the same solid support. The molecular label (311) may be different for at least two or more oligonucleotides on the solid support. The solid support may comprise 100 or more oligonucleotides. The solid support may comprise 1000 or more oligonucleotides. The solid support may comprise 10000 or more oligonucleotides. The solid support may comprise 100000 or more oligonucleotides.

[00247] In addition to a universal label, a cellular label, and a molecular label, an oligonucleotide may comprise a target binding region. A target binding region may comprise a nucleic acid sequence that may bind to a target nucleic acid (*e.g.*, a cellular nucleic acid to be analyzed). A target binding region may be a gene specific sequence. For example, a target binding region may comprise a nucleic acid sequence that may attach (*e.g.*, hybridize) to a specific location of a specific target nucleic acid. A target binding region may comprise a non-specific target nucleic acid sequence. A non-specific target nucleic acid sequence may refer to a sequence that may bind to multiple target nucleic acids, independent of the specific sequence of the target nucleic acid. For example, target binding region may comprise a random multimer sequence or an oligo dT sequence (*e.g.*, a stretch of thymidine nucleotides that may hybridize to a poly-adenylation tail on mRNAs). A random multimer sequence can be, for example, a random dimer, trimer, quatramer, pentamer, hexamer, septamer, octamer, nonamer, decamer, or higher multimer sequence of any length. A target binding region may be at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A target binding region may be at most about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length.

[00248] An oligonucleotide may comprise a plurality of labels. For example an oligonucleotide may comprise at least about 1, 2, 3, 4, 5, 6, 7, or 8 or more universal labels. An oligonucleotide may comprise at most about 1, 2, 3, 4, 5, 6, 7, or 8 or more universal labels. An oligonucleotide may comprise at least about 1, 2, 3, 4, 5, 6, 7, or 8 or more cellular labels. An oligonucleotide may comprise at most about 1, 2, 3, 4, 5, 6, 7, or 8 or more cellular labels. An oligonucleotide may comprise at least about 1, 2, 3, 4, 5, 6, 7, or 8 or more molecular labels. An oligonucleotide may comprise at most about 1, 2, 3, 4, 5, 6, 7, or 8 or more molecular labels. An oligonucleotide may comprise at least about 1, 2, 3, 4, 5, 6, 7, or 8 or more sample labels. An oligonucleotide may comprise at most about 1, 2, 3, 4, 5, 6, 7, or 8 or more sample labels. An oligonucleotide may comprise at least about 1, 2, 3, 4, 5, 6, 7, or 8

IDT Exhibit 1018.0929

or more target binding regions. An oligonucleotide may comprise at most about 1, 2, 3, 4, 5, 6, 7, or 8 or more target binding regions.

[00249] When an oligonucleotide comprises more than one of a type of label (*e.g.*, more than one cellular label or more than one molecular label), the labels may be interspersed with a linker label sequence. A linker label sequence may be at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A linker label sequence may be at most about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. In some instances, a linker label sequence is 12 nucleotides in length. A linker label sequence may be used to facilitate the synthesis of the oligonucleotide, such as diagrammed in **FIG. 2A**.

[00250] The number of oligonucleotides conjugated to a solid support may be 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10-fold more than the number of target nucleic acids in a cell. In some instances, at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100% of the oligonucleotides are bound by a target nucleic acid. In some instances, at most 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100% of the oligonucleotides are bound by a target nucleic acid. In some instances, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 or more different target nucleic acids are captured by the oligonucleotides on a solid support. In some instances, at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 or more different target nucleic acids are captured by the oligonucleotides on a solid support.

[00251] A polymer may comprise additional solid supports. For example, a polymer may be dotted with beads. The beads may be spatially located at different regions of the polymer. The beads or supports comprising oligonucleotides of the disclosure may be spatially addressed. The beads or supports may comprise a barcode corresponding to a spatial address on the polymer. For example, each bead or support of a plurality of beads or supports may comprise barcode that corresponds to a position on a polymer, such as a position on an array or a particular microwall of a plurality of microwells. The spatial address can be decoded to determine the location from which a bead or support was positioned. For example, a spatial address, such as a barcode, can be decoded by hybridization of an oligonucleotide to the barcode or by sequencing the barcode. Alternatively, beads or supports can bear other types of barcodes, such as graphical features, chemical groups, colors, fluorescence, or combinations any combination thereof, for spatial address decoding purposes.

[00252] The methods and kits disclosed herein may comprise one or more sets of molecular barcodes. One or more molecular barcodes may comprise a sample index region and a label region. Two or more molecular barcodes of a set of molecular barcodes may comprise the same sample index region and two or more different label regions. Two or more molecular

IDT Exhibit 1018.0930

barcodes of two or more sets of molecular barcodes may comprise two or more different sample index regions. Two or more molecular barcodes from a set of molecular barcodes may comprise different label regions. Two or more molecular barcodes of two or more sets of molecular barcodes may comprise the same label region. Molecular barcodes from two or more sets of molecular barcodes may differ by their sample index regions. Molecular barcodes from two or more sets of molecular barcodes may be similar based on their label regions.

[00253] The molecular barcodes may further comprise a target specific region, an adapter region, a universal PCR region, a target specific region or any combination thereof. The molecular barcode may comprise a universal PCR region and a target specific region. The molecular barcode may comprise one or more secondary structures. The molecular barcode may comprise a hairpin structure. The molecular barcode may comprise a target specific region and a cleavable stem.

[00254] The methods and kits disclosed herein may comprise one or more sets of sample tags. One or more sample tags may comprise a sample index region. One or more sample tags may comprise a sample index region. Two or more sample tags of a set of sample tags may comprise the same sample index region. Two or more sample tags of two or more sets of sample tags may comprise two or more different sample index regions.

[00255] The sample tags may further comprise a target specific region, an adapter region, a universal PCR region, a target specific region or any combination thereof. The sample tag may comprise a universal PCR region and a target specific region. The sample tag may comprise one or more secondary structures. The sample tag may comprise a hairpin structure. The sample tag may comprise a target specific region and a cleavable stem.

[00256] The methods and kits disclosed herein may comprise one or more sets of, molecular identifier labels. One or more molecular identifier labels may comprise a label region. One or more molecular identifier labels may comprise a label region. Two or more molecular identifier labels of a set of molecular identifier labels may comprise two or more different label regions. Two or more molecular identifier labels of two or more sets of molecular identifier labels may comprise two or more identical label regions. The molecular identifier labels may further comprise a target specific region, an adapter region, a universal PCR region, a target specific region or any combination thereof. The molecular identifier label may comprise a universal PCR region and a target specific region. The molecular identifier label may comprise one or more secondary structures. The molecular identifier label may

IDT Exhibit 1018.0931

comprise a hairpin structure. The molecular identifier label may comprise a target specific region and a cleavable stem.

[00257] The molecular barcode, sample tag or molecular identifier label may comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or base pairs. In another example, the sample tag or molecular identifier label comprises at least about 1500, 2,000; 2500, 3,000; 3500, 4,000; 4500, 5,000; 5500, 6,000; 6500, 7,000; 7500, 8,000; 8500, 9,000; 9500, or 10,000 nucleotides or base pairs.

[00258] The molecular barcodes, sample tags or molecular identifier labels may be multimers, *e.g.,* random multimers. A multimer sequence can be, for example, a non-random or random dimer, trimer, quatramer, pentamer, hexamer, septamer, octamer, nonamer, decamer, or higher multimer sequence of any length. The tags may be randomly generated from a set of mononucleotides. The tags may be assembled by randomly incorporating mononucleotides.

[00259] The molecular barcodes, sample tags or molecular identifier labels may also be assembled without randomness, to generate a library of different tags which are not randomly generated but which includes sufficient numbers of different tags to practice the methods.

[00260] In some embodiments a molecular barcode, sample tag or molecular identifier label may comprise a cutback in a target nucleic acid. The cutback may be, for example, an enzymatic digestion of one or both ends of a target nucleic acid. The cutback may be used in conjunction with the addition of added molecular barcodes, sample tags (*e.g.*, sample index region, sample label), cellular label, and molecular identifier labels (*e.g.*, molecular label). The combination of the cutback and the added tags may contain information related to the particular starting molecule. By adding a random cutback to the molecular barcode, sample tag or molecular identifier label, a smaller diversity of the added tags may be necessary for counting the number of target nucleic acids when detection allows a determination of both the random cutback and the added oligonucleotides.

[00261] The molecular barcode, sample tag or molecular identifier label may comprise a target specific region. The target specific region may comprise a sequence that is complementary to the molecule. In some instances, the molecule is an mRNA molecule and the target specific region comprises an oligodT sequence that is complementary to the polyA tail of the mRNA molecule. The target specific region may also act as a primer for DNA and/or RNA synthesis. For example, the oligodT sequence of the target specific region may act as a primer for first strand synthesis of a cDNA copy of the mRNA molecule.

IDT Exhibit 1018.0932

Alternatively, the target specific region comprises a sequence that is complementary to any portion of the molecule. In other instances, the target specific region comprises a random sequence that may be hybridized or ligated to the molecule. The target specific region may enable attachment of the sample tag or molecular identifier label to the molecule. Attachment of the sample tag or molecular identifier label may occur by any of the methods disclosed herein (*e.g.*, hybridization, ligation). In some instances, the target specific region comprises a sequence that is recognized by one or more restriction enzymes. The target specific region may comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or base pairs. In another example, the target specific region comprises at least about 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, or 10000 nucleotides or base pairs. Preferably, the target specific region comprises at least about 5-10, 10-15, 10-20, 10-30, 15-30, or 20-30 nucleotides or base pairs.

[00262] In some instances, the target specific region is specific for a particular gene or gene product. For example, the target specific region comprises a sequence complementary to a region of a p53 gene or gene product. Therefore, the sample tags and molecular identifier labels may only attach to molecules comprising the p53-specific sequence. Alternatively, the target specific region is specific for a plurality of different genes or gene products. For example, the target specific region comprises an oligodT sequence. Therefore, the sample tags and molecular identifier labels may attach to any molecule comprising a polyA sequence. In another example, the target specific region comprises a random sequence that is complementary to a plurality of different genes or gene products. Thus, the sample tag or molecular identifier label may attach to any molecule with a sequence that is complementary to the target specific region. In other instances, the target specific region comprises a restriction site overhang (*e.g.*, EcoRI sticky-end overhang). The sample tag or molecular identifier label may ligate to any molecule comprising a sequence complementary to the restriction site overhang.

[00263] In some instances, the target specific region is specific for a particular microRNA or microRNA product. For example, the target specific region comprises a sequence complementary to a region of a specific microRNA or microRNA product. For example, the target specific regions comprise sequences complementary to regions of a specific panel of microRNAs or panel of microRNA products. Therefore, the sample tags and molecular identifier labels may only attach to molecules comprising the micoRNA-specific sequence. Alternatively, the target specific region is specific for a plurality of different micoRNAs or

IDT Exhibit 1018.0933

micoRNA products. For example, the target specific region comprises a sequence complimentary to a region comprised in two or more microRNAs, such as a panel of microRNAs containing a common sequence. Therefore, the sample tags and molecular identifier labels may attach to any molecule comprising the common microRNA sequence. In another example, the target specific region comprises a random sequence that is complementary to a plurality of different microRNAs or microRNA products. Thus, the sample tag or molecular identifier label may attach to any microRNA molecule with a sequence that is complementary to the target specific region. In other instances, the target specific region comprises a restriction site overhang (*e.g.*, EcoRI sticky-end overhang). The sample tag or molecular identifier label may ligate to any microRNA molecule comprising a sequence complementary to the restriction site overhang.

[00264] The molecular barcode or molecular identifier label disclosed herein often comprises a label region. The label region may be used to uniquely identify occurrences of target species thereby marking each species with an identifier that may be used to distinguish between two otherwise identical or nearly identical targets. The label region of the plurality of sample tags and molecular identifier labels may comprise a collection of different semiconductor nanocrystals, metal compounds, peptides, oligonucleotides, antibodies, small molecules, isotopes, particles or structures having different shapes, colors, barcodes or diffraction patterns associated therewith or embedded therein, strings of numbers, random fragments of proteins or nucleic acids, different isotopes, or any combination thereof. The label region may comprise a degenerative sequence. The label region may comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or base pairs. In another example, the label region comprises at least about 1500; 2,000; 2500, 3,000; 3500, 4,000; 4500, 5,000; 5500, 6,000; 6500, 7,000; 7500, 8,000; 8500, 9,000; 9500, or 10,000 nucleotides or base pairs. Preferably, the label region comprises at least about 10-30, 15-40, or 20-50 nucleotides or base pairs.

[00265] In some instances, the molecular barcode, sample tag or molecular identifier label comprises a universal primer binding site. The universal primer binding site allows the attachment of a universal primer to the labeled-molecule and/or labeled-amplicon. Universal primers are well known in the art and include, but are not limited to, -47F (M13F), alfaMF, AOX3', AOX5', BGH_r, CMV_-30, CMV_-50, CVM_f, LACrmt, lamgda gt10F, lambda gt 10R, lambda gt11F, lambda gt11R, M13 rev, M13Forward(-20), M13Reverse, male, p10SEQP_pQE, pA_-120, pet_4, pGAP Forward, pGL_RVpr3, pGLpr2_R, pKLAC1_4,

IDT Exhibit 1018.0934

pQE_FS, pQE_RS, puc_U1, puc_U2, revers_A, seq_IRES_tam, seq_IRES_zpet, seq_ori, seq_PCR, seq_pIRES-, seq_pIRES+, seq_pSecTag, seq_pSecTag+, seq_retro+PSI, SP6, T3-prom, T7-prom, and T7-term_Inv. Attachment of the universal primer to the universal primer binding site may be used for amplification, detection, and/or sequencing of the labeled-molecule and/or labeled-amplicon. The universal primer binding site may comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or base pairs. In another example, the universal primer binding site comprises at least about 1500; 2,000; 2500, 3,000; 3500, 4,000; 4500, 5,000; 5500, 6,000; 6500, 7,000; 7500, 8,000; 8500, 9,000; 9500, or 10,000 nucleotides or base pairs. Preferably, the universal primer binding site comprises 10-30 nucleotides or base pairs.

[00266] The molecular barcode, sample tag or molecular identifier label may comprise an adapter region. The adapter region may enable hybridization of one or more probes. The adapter region may enable hybridization of one or more HCR probes.

[00267] The molecular barcode, sample tag or molecular identifier label may comprise one or more detectable labels.

[00268] The molecular barcode, sample tag or molecular identifier label may act as an initiator for a hybridization chain reaction (HCR). The adapter region of the sample tag or molecular identifier label may act as an initiation for HCR. The universal primer binding site may act as an initiator for HCR.

[00269] In some instances, the molecular barcode, sample tag or molecular identifier label is single-stranded. In other instances, the molecular barcode, sample tag or molecular identifier label is double-stranded. The molecular barcode, sample tag or molecular identifier label may be linear. Alternatively, the molecular barcode, sample tag or molecular identifier label comprises a secondary structure. As used herein, "secondary structure" includes tertiary, quaternary, etc…structures. In some instances, the secondary structure is a hairpin, a stem-loop structure, an internal loop, a bulge loop, a branched structure or a pseudoknot, multiple stem loop structures, cloverleaf type structures or any three dimensional structure. In some instances, the secondary structure is a hairpin. The hairpin may comprise an overhang sequence. The overhang sequence of the hairpin may act as a primer for a polymerase chain reaction and/or reverse transcription reaction. The overhang sequence comprises a sequence that is complementary to the molecule to which the sample tag or molecular identifier label is attached and the overhang sequence hybridizes to the molecule. The overhang sequence may be ligated to the molecule and acts as a template for a polymerase chain reaction and/or

IDT Exhibit 1018.0935