

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

INTEGRATED DNA TECHNOLOGIES, INC.,
Petitioner,

v.

TECAN GENOMICS, INC.,
Patent Owner.

IPR2024-01504

U.S. Patent No. 11,098,357 B2

PETITIONER'S DEMONSTRATIVE EXHIBITS

***United States Patent and Trademark Office
Before the Patent Trial and Appeal Board***

***Integrated DNA Technologies, Inc.
v. Tecan Genomics, Inc.***

Petition for Inter Partes Review
Case No. IPR2024-01504
U.S. Patent No. 11,098,357

Petitioner's Demonstratives

Oral Hearing: January 22, 2026

Claims 1 and 2

'357 Patent



US011098357B2

(12) **United States Patent**
Amorese et al.

(10) **Patent No.:** US 11,098,357 B2
(45) **Date of Patent:** *Aug. 24, 2021

(54) **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

(71) **Applicant:** NUGEN TECHNOLOGIES, INC., Redwood City, CA (US)

(72) **Inventors:** Douglas A. Amorese, Los Altos, CA (US); Jonathan Seolnick, San Francisco, CA (US); Ben Schroeder, San Mateo, CA (US)

(73) **Assignee:** TECAN GENOMICS, INC., Redwood City, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.
This patent is subject to a terminal disclaimer.

(21) **Appl. No.:** 16/743,604

(22) **Filed:** Jan. 15, 2020

(65) **Prior Publication Data**

US 2020/0239953 A1 Jul. 30, 2020

Related U.S. Application Data

(66) Continuation of application No. 15/406,002, filed on Jan. 13, 2017, now Pat. No. 10,570,448, which is a division of application No. 14/540,917, filed on Nov. 13, 2014, now Pat. No. 9,546,399.

(66) Provisional application No. 61/903,826, filed on Nov. 13, 2013.

(51) **Int. Cl.**

C12Q 1/6874 (2018.01)

C12Q 1/6869 (2018.01)

(52) **U.S. Cl.**

C12Q 1/6874 (2013.01); **C12Q 1/6869** (2013.01); **C12Q 2525/191** (2013.01); **C12Q 2525/192** (2013.01)

(58) **Field of Classification Search**

CPC **C12Q 1/6869**; **C12Q 2525/191**; **C12Q 2525/192**

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,362,867 A 12/1982 Paddock
4,458,066 A 7/1984 Caruthers et al.
4,469,863 A 9/1984 Ts'o et al.
4,582,877 A 4/1986 Fairchok et al.
4,876,187 A 10/1989 Duck et al.
4,923,063 A 5/1990 Golias
4,935,557 A 6/1990 Seybold
4,942,124 A 7/1990 Church
4,988,617 A 1/1991 Landegren et al.
4,996,143 A 2/1991 Haller et al.
5,011,769 A 4/1991 Duck et al.
5,034,506 A 7/1991 Summerton et al.

5,035,996 A 7/1991 Hartley

5,043,272 A 8/1991 Hartley

5,082,830 A 1/1992 Patel et al.

5,090,591 A 2/1992 Ling

5,130,238 A 7/1992 Malik et al.

5,169,766 A 12/1992 Schuster et al.

5,171,534 A 12/1992 Smith et al.

5,194,570 A 3/1993 Deming et al.

5,216,141 A 6/1993 Benner

5,234,409 A 8/1993 Jovin et al.

5,235,033 A 8/1993 Summerton et al.

5,242,794 A 9/1993 Whitley et al.

5,284,242 A 1/1993 Oltus

5,386,023 A 1/1995 Sanghvi et al.

5,399,491 A 3/1995 Kacian et al.

5,409,818 A 4/1995 Darvey et al.

5,418,149 A 5/1995 Cellular et al.

5,422,271 A 6/1995 Chen et al.

5,427,929 A 6/1995 Richards et al.

5,480,784 A 1/1996 Kacian et al.

5,494,810 A 2/1996 Darvey et al.

5,508,169 A 4/1996 Duggan et al.

5,508,178 A 4/1996 Rose et al.

5,510,270 A 4/1996 Foster et al.

5,525,471 A 6/1996 Zeng

5,545,512 A 8/1996 Van Gelder et al.

5,554,516 A 9/1996 Kacian et al.

5,554,517 A 9/1996 Darvey et al.

5,556,752 A 9/1996 Lockhart et al.

5,565,540 A 10/1996 Cheschik et al.

5,573,913 A 11/1996 Rasmussen et al.

5,578,832 A 11/1996 Tridion et al.

5,580,539 A 12/1996 Hampton et al.

5,602,240 A 2/1997 De Meester et al.

(Continued)

OTHER PUBLICATIONS

Illumina publication No. 770-2008-011, Multiplexed Sequencing with the Illumina Genome Analyzer System, pp. 1-4 (Year 2008) *

Vaidi et al., Next generation sequencing for molecular diagnosis of neurovascular diseases, *Acta Neuropathol.*, vol. 126, pp. 275-285 plus supplemental material pp. 1-10 (Year 2012) *

Amos, 2000, DNA pooling in mutation detection with reference to sequence analysis, *Am J Hum Genet* 66:1689-1692.

Belloso, 2014, cnaCapSeq: detecting copy number variation in long-range targeted resequencing data, *Nucleic Acids Res* 42:201-158.

Benson, 2013, Genbank, *Nucl Acids Res* 41:D36-D42.

Bloomquist, 2013, Targeted RNA Sequencing with Competitive Multiplex-PCR, *Amplification Libraries*, PLOS ONE 8(11):e79120.

Bloch, 2013, Comparison of commercially available target enrichment methods for next-generation sequencing, *J Biomolecular Tech* 24:73-86.

(Continued)

Primary Examiner — Teresa E Strezelska

(74) **Attorney, Agent, or Firm** — Brown Rudnick LLP; Thomas C. Meyers

(57) **ABSTRACT**

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed.

10 Claims, 6 Drawing Sheets

Specification includes a Sequence Listing.

IDT Exhibit 1004.0001

1. A method for detecting duplicate sequencing reads, the method comprising:
ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;
amplifying the adaptor-ligated fragments into amplicons;
sequencing the amplicons to produce sequence reads that include identifier and target sequences; and
identifying sequence reads with identical identifier and target sequences as duplicates.
2. The method of claim 1, further comprising removing a duplicate read from the sequence reads.

Claims 1 and 2

The Physical Limitations

1. A method for detecting duplicate sequencing reads, the method comprising:
ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;
amplifying the adaptor-ligated fragments into amplicons;
sequencing the amplicons to produce sequence reads that include identifier and target sequences; and

The Computer-Aided Limitations

identifying sequence reads with identical identifier and target sequences as duplicates.

2. The method of claim 1, further comprising removing a duplicate read from the sequence reads.

A circular portrait of Dr. David A. Clark, a man with short brown hair and glasses, smiling. He is wearing a grey shirt. The portrait is set against a white background and is framed by a thin white border.

Dr. Peter Sims

Expert for IDT

Pet. at 24-70; EX1079 ¶¶ 343-360, 367-393, 563-576, 584-609

Computer-Aided Analysis of Sequencing Data Was Well-Known



**Dr. Carlos
Bustamante**
Expert for Tecan

- 1) Read Quality Control and Trimming,
- 2) Reference Genome Alignment,
- 3) Format conversion and coordinate sorting/indexing (SAM/BAM),
- 4) Sample Demultiplexing,
- 5) Duplicate Identification,
- 6) Recalibration and Realignment, and
- 7) Variant/Feature Calling

Petitioner's Hearing Plan

1 Claim Construction

2 Anticipation – Bielas (EX1013)

3 Obviousness – lafrate & Kivioja (EX1011 & 1012)

Petitioner's Hearing Plan

1 Claim Construction

2 Anticipation – Bielas (EX1013)

3 Obviousness – Iafrate & Kivioja (EX1011 & 1012)

Construing the Computer-Aided Limitation in Claim 1

'357 Patent



US011098357B2

(12) **United States Patent**
Amorese et al.

(10) **Patent No.:** US 11,098,357 B2
(45) **Date of Patent:** *Aug. 24, 2021

(54) **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

(71) **Applicant:** NUGEN TECHNOLOGIES, INC., Redwood City, CA (US)

(72) **Inventors:** Douglas A. Amorese, Los Altos, CA (US); Jonathan Seolnick, San Francisco, CA (US); Ben Schroeder, San Mateo, CA (US)

(73) **Assignee:** TECAN GENOMICS, INC., Redwood City, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days. This patent is subject to a terminal disclaimer.

(21) **Appl. No.:** 16/743,604

(22) **Filed:** Jan. 15, 2020

(65) **Prior Publication Data**

US 2020/0239953 A1 Jul. 30, 2020

Related U.S. Application Data

(66) Continuation of application No. 15/406,002, filed on Jan. 13, 2017, now Pat. No. 10,570,448, which is a division of application No. 14/540,917, filed on Nov. 13, 2014, now Pat. No. 9,546,399.
(67) Provisional application No. 61/903,826, filed on Nov. 13, 2013.

(51) **Int. Cl.**
C12Q 1/6874 (2018.01)
C12Q 1/6869 (2018.01)
U.S. Cl.
CPC **C12Q 1/6874** (2013.01); **C12Q 1/6869** (2013.01); **C12Q 2525/191** (2013.01); **C12Q 2525/122** (2013.01)

(58) **Field of Classification Search**
CPC **C12Q 1/6869**; **C12Q 2525/191**; **C12Q 2525/122**
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,362,867 A 12/1982 Paddock
4,458,066 A 7/1984 Caruthers et al.
4,469,863 A 9/1984 Te'o et al.
4,582,877 A 4/1986 Fairchok et al.
4,876,187 A 10/1989 Duck et al.
4,923,063 A 5/1990 Golias
4,935,557 A 6/1990 Seydahl
4,942,124 A 7/1990 Church
4,988,617 A 1/1991 Landegren et al.
4,996,143 A 2/1991 Haller et al.
5,011,769 A 4/1991 Duck et al.
5,034,506 A 7/1991 Summerton et al.

5,035,996 A 7/1991 Hartley
5,043,272 A 8/1991 Hartley
5,082,830 A 1/1992 Bialski et al.
5,090,591 A 2/1992 Long
5,130,238 A 7/1992 Malik et al.
5,169,766 A 12/1992 Schuster et al.
5,171,534 A 12/1992 Smith et al.
5,194,370 A 3/1993 Denzinger et al.
5,216,141 A 6/1993 Renner
5,234,409 A 8/1993 Biron et al.
5,235,033 A 8/1993 Summerton et al.
5,242,794 A 9/1993 Whitley et al.
5,284,242 A 1/1993 Orlan
5,386,023 A 1/1995 Sanghvi et al.
5,399,491 A 3/1995 Kacian et al.
5,409,818 A 4/1995 Darvey et al.
5,418,149 A 5/1995 Cellinal et al.
5,422,271 A 6/1995 Chen et al.
5,427,929 A 6/1995 Richards et al.
5,480,784 A 1/1996 Kacian et al.
5,494,810 A 2/1996 Darvey et al.
5,508,169 A 4/1996 Dragan et al.
5,508,178 A 4/1996 Rose et al.
5,510,270 A 4/1996 Foster et al.
5,525,471 A 6/1996 Zeng
5,545,522 A 8/1996 Van Gelder et al.
5,554,216 A 9/1996 Kacian et al.
5,554,517 A 9/1996 Darvey et al.
5,556,752 A 9/1996 Lockhart et al.
5,565,540 A 10/1996 Cheschik et al.
5,573,913 A 11/1996 Roumney et al.
5,578,832 A 11/1996 Trudon et al.
5,580,339 A 12/1996 Hampton et al.
5,602,240 A 2/1997 De Meester et al.
(Continued)

FOREIGN PATENT DOCUMENTS

CA 2444926 A1 11/2002
CN 1661102 A 8/2005
(Continued)

OTHER PUBLICATIONS

Illumina publication No. 770-2008-011, Multiplexed Sequencing with the Illumina Genome Analyzer System, pp. 1-4 (Year: 2008) *
Vaidi et al., Next generation sequencing for molecular diagnosis of neurovascular diseases, *Acta Neuropathol.*, vol. 126, pp. 275-285 plus supplemental material pp. 1-10 (Year: 2012) *
Amos, 2009, DNA pooling in mutation detection with reference to sequence analysis, *Am J Hum Genet* 66:1689-1692.
Bellos, 2014, *casCapSeq*: detecting copy number variation in long-range targeted resequencing data, *Nucleic Acids Res* 42:201-158.
Benson, 2013, Genbank, *Nucl Acids Res* 41:D36-D42.
Blomquist, 2013, Targeted RNA Sequencing with Competitive Multiplex-PCR, *Amplification Libraries*, PLOS ONE 8(11):e79120.
Desh, 2013, Comparison of commercially available target enrichment methods for next-generation sequencing, *J Biomolecular Tech* 24:73-86.
(Continued)

Primary Examiner — Teresa E Strezelska

(74) **Attorney, Agent, or Firm** — Brown Rudnick LLP;
Thomas C. Meyers

(57) **ABSTRACT**

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed.

10 Claims, 6 Drawing Sheets

Specification includes a Sequence Listing.

IDT Exhibit 1004.0001

1. A method for detecting duplicate sequencing reads, the method comprising:
ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;
amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences; and
identifying sequence reads with identical identifier and target sequences as duplicates.

There Is No Dispute Over the Meaning of “Duplicates”

Petitioner’s Reply

A. “Duplicates” Require 100% Identity Between Two Sequences.

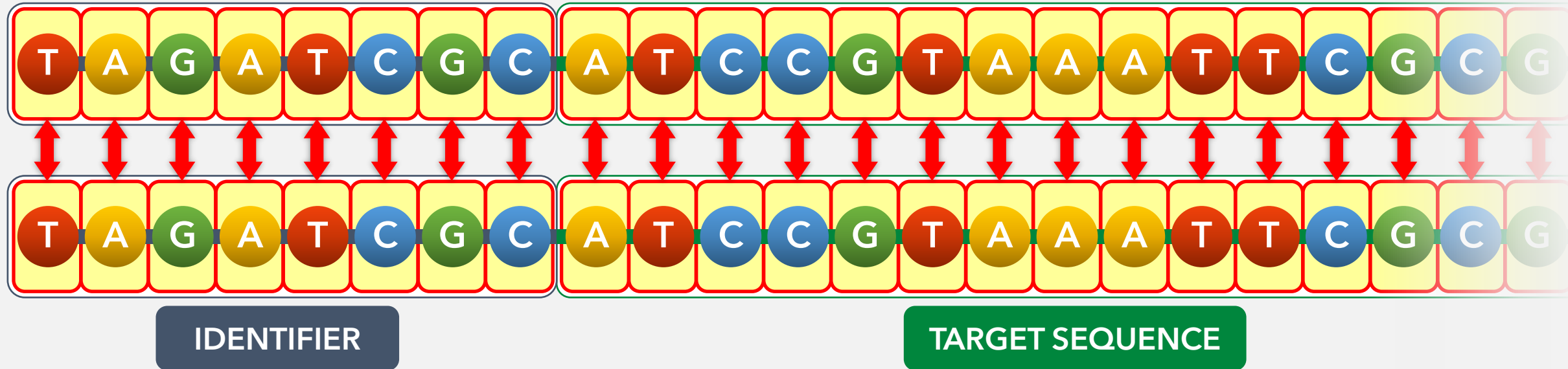
Claim 1 provides that “duplicates” are “sequence reads with *identical* identifier and target sequences.” EX1004 at 25:46-47 (emphasis added). There is no dispute that the term “duplicates” as used in the patent requires the sequences to be identical. See POR at 16-17; EX1124 at 31:1-38:3 (discussing demonstratives EX1113, EX1114, and EX1115). Requiring a “duplicate” to be 100% identical, i.e., having a Hamming distance of zero, EX1124 at 37:24-38:3, accords with the plain and customary meaning of “identical” as used in the claim. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312-13 (Fed. Cir. 2005) (en banc).

Tecan’s Sur-Reply

B. A Method “Comprising” a Step that Identifies “Duplicates”

There appears to be no dispute that two sequences are duplicates when the reads being compared are identical. When read in combination with the “comprising” transitional phrase of the claims, the claims only require that duplicates are identified. The claims are agnostic as to whether the method also identifies non-duplicates, including those sequences that would be captured with a Hamming distance greater than zero and potential errors.

"Duplicates" Requires 100% Sequence Identity



comprises the target sequence. In some embodiments, a sequencing read comprises a target sequence and an identifier sequence. For example, the target sequence is sequenced with the identifier site. In some embodiments, the

The Computer-Aided Limitation Is Straightforward

Claim Term

“identifying sequence reads with identical identifier and target sequences as duplicates”

'357 Patent, Claim 1

Tecan Adds Two Unsupported Limitations

Claim Term	Tecan's Construction
“identifying sequence reads with identical identifier and target sequences as duplicates” '357 Patent, Claim 1	“identifying sequence reads as duplicates through a <u>step</u> that uses both the ~~with identical~~ identifier and <u>at least a portion of the</u> target sequences ~~as~~ duplicates”

Tecan Imposes Simultaneity With Its “Step” Limitation

Tecan’s Sur-Reply

The dispute here is whether the identifying step of the claims allows for (1) performing a series of steps that eventually lead to the identification of duplicates, without regard as to how the identifier and target sequences are used, as IDT contends, or instead requires (2) using the identifier and target sequences in combination in the step that identifies duplicates, as Tecan contends. This is not a question of the order of steps, but a question of what is required within the identification step. Here, the prosecution history, which IDT largely ignores,

* * *

There is no ambiguity in Applicants’ representations that the identification of duplicates requires that both the target sequence and the identifier be used together.

The Board Should Adopt the Plain and Ordinary Meaning of the Computer-Aided Limitation

Claim Term	Tecan's Construction	Petitioner's Construction
“identifying sequence reads with identical identifier and target sequences as duplicates” '357 Patent, Claim 1	“identifying sequence reads as duplicates through <u>a step</u> that uses both the ~~with identical~~ identifier and <u>at least a portion of the</u> target sequences ~~as duplicates~~”	“identifying sequence reads with identical identifier and target sequences as duplicates.” This limitation requires the entire “target sequence” and does not require a single “step.”

Simultaneity Is Unsupported by the Claim Language

'357 Patent



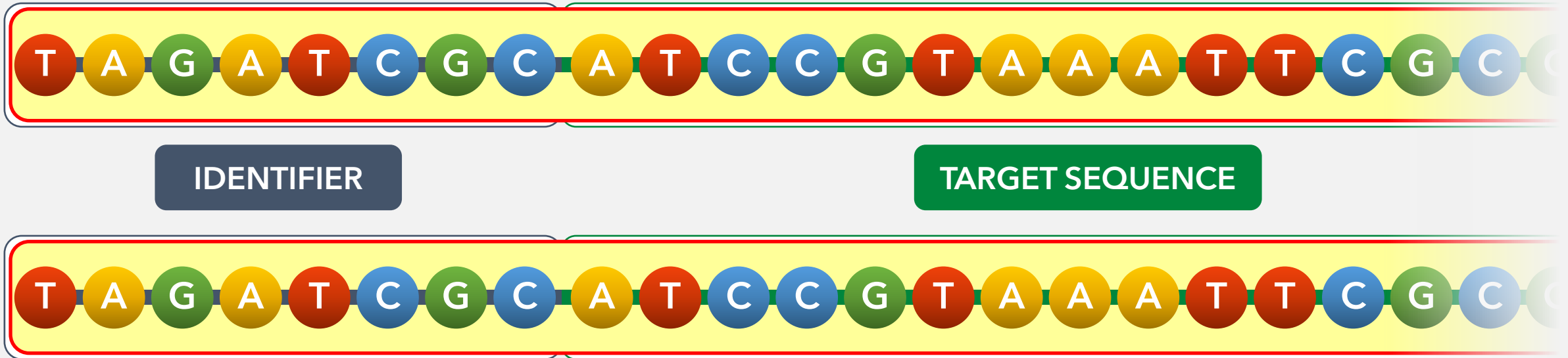
US011098357B2

(12) United States Patent Amorese et al.	(10) Patent No.: US 11,098,357 B2 (45) Date of Patent: *Aug. 24, 2021
(54) COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	5,035,996 A 7/1991 Hartley 5,043,272 A 8/1991 Hartley 5,082,830 A 1/1992 Bartel et al. 5,090,591 A 2/1992 Long 5,130,238 A 7/1992 Malik et al. 5,169,766 A 12/1992 Schuster et al. 5,171,534 A 12/1992 Smith et al. 5,194,570 A 3/1993 Denzinger et al. 5,216,141 A 6/1993 Benner 5,234,409 A 8/1993 Jovin et al. 5,235,033 A 8/1993 Summerton et al. 5,242,794 A 9/1993 Whitley et al. 5,284,242 A 1/1993 Orlow 5,386,023 A 1/1995 Sanghvi et al. 5,399,491 A 3/1995 Kacian et al. 5,409,818 A 4/1995 Darvey et al. 5,418,149 A 5/1995 Cellular et al. 5,422,271 A 6/1995 Chen et al. 5,427,929 A 6/1995 Richards et al. 5,480,781 A 1/1996 Kacian et al. 5,494,810 A 2/1996 Darvey et al. 5,508,169 A 4/1996 Dragan et al. 5,508,178 A 4/1996 Rose et al. 5,510,270 A 4/1996 Foster et al. 5,525,471 A 6/1996 Zeng 5,545,512 A 8/1996 Van Gelder et al. 5,554,516 A 9/1996 Kacian et al. 5,554,517 A 9/1996 Darvey et al. 5,556,752 A 9/1996 Lockhart et al. 5,565,540 A 10/1996 Cheschik et al. 5,573,913 A 11/1996 Roumney et al. 5,578,832 A 11/1996 Tridion et al. 5,580,539 A 12/1996 Hampton et al. 5,602,240 A 2/1997 De Moenneker et al. (Continued)
(71) Applicant: NUGEN TECHNOLOGIES, INC. , Redwood City, CA (US)	
(72) Inventors: Douglas A. Amorese , Los Altos, CA (US); Jonathan Seolnick , San Francisco, CA (US); Ben Schroeder , San Mateo, CA (US)	
(73) Assignee: TECAN GENOMICS, INC. , Redwood City, CA (US)	
(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days. This patent is subject to a terminal dis- claimer.	
(21) Appl. No.: 16,743,604	
(22) Filed: Jan. 15, 2020	
(65) Prior Publication Data US 2020/023953 A1 Jul. 30, 2020	
Related U.S. Application Data	
(60) Continuation of application No. 15/406,002, filed on Jan. 13, 2017, now Pat. No. 10,570,448, which is a division of application No. 14/540,917, filed on Nov. 13, 2014, now Pat. No. 9,546,399.	
(60) Provisional application No. 61/903,826, filed on Nov. 13, 2013.	
(51) Int. Cl. C12Q 1/6874 (2018.01) C12Q 1/6869 (2018.01)	
(52) U.S. Cl. CPC ——— C12Q 1/6874 (2013.01); C12Q 1/6869 (2013.01); C12Q 2525/179 (2013.01); C12Q 2525/191 (2013.01)	
(58) Field of Classification Search CPC ——— C12Q 1/6869 ; C12Q 2525/191 ; C12Q 2525/122 See application file for complete search history.	
(56) References Cited U.S. PATENT DOCUMENTS 4,362,867 A 12/1982 Paddock 4,458,066 A 7/1984 Caruthers et al. 4,469,863 A 9/1984 Ts'o et al. 4,582,877 A 4/1986 Fairchok et al. 4,876,187 A 10/1989 Duck et al. 4,923,063 A 5/1990 Golias 4,935,557 A 6/1990 Seybold 4,942,124 A 7/1990 Church 4,988,617 A 1/1991 Landegren et al. 4,996,143 A 2/1991 Haller et al. 5,011,769 A 4/1991 Duck et al. 5,034,506 A 7/1991 Summerton et al.	

IDT Exhibit 1004.0001

1. A method for detecting duplicate sequencing reads, the method comprising:
ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;
amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences; and identifying sequence reads with identical identifier and target sequences as duplicates.

Tecan's Construction Requires Simultaneity



Tecan's Sur-Reply

There is no ambiguity in Applicants' representations that the identification of duplicates requires that both the target sequence and the identifier be used together.

The Preceding Limitations Do Not Support Simultaneity

Tecan's Response

Starting with the claim, it is clear from the claim language itself that Applicants intended that the claimed method included an independent step of identifying duplicates by using both the identifier and at least a portion of the target sequence. Each of the steps of the claimed method include specific elements necessary for that step, and the identifying step is no different. An interpretation of this term that allows the identification of duplicates with only one of the claimed elements essentially reads out the rest of this limitation. For example, if this claim

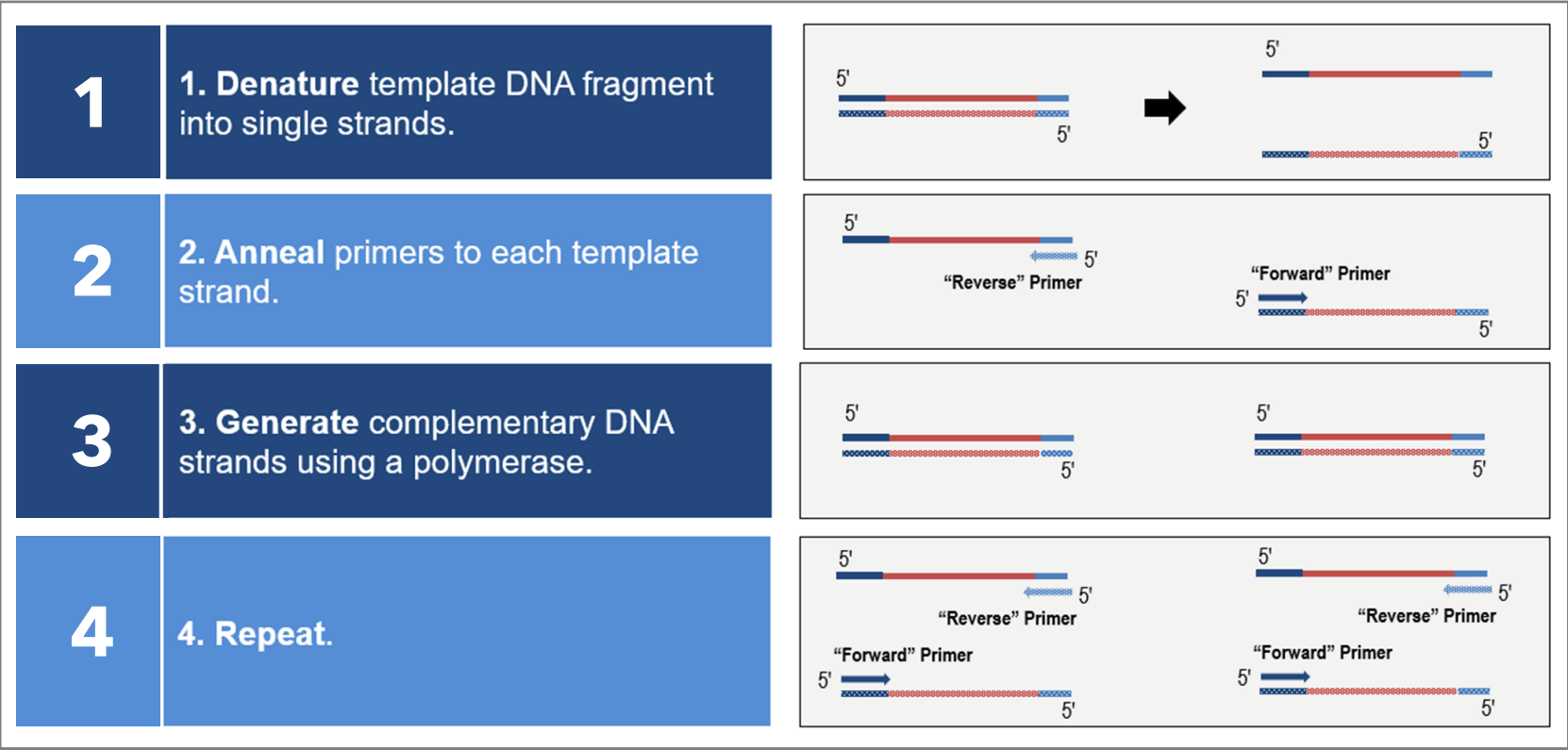
Claim 1

1. A method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;
amplifying the adaptor-ligated fragments into amplicons;
sequencing the amplicons to produce sequence reads that include identifier and target sequences; and
identifying sequence reads with identical identifier and target sequences as duplicates.

The Amplifying Limitation in Claim 1 Is a *Multistep* Process

The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as for amplification of a target sequence (e.g., in PCR). The



Imposing Simultaneity Inappropriately Departs From the Plain and Ordinary Meaning

Apple Inc. v. Gesture Tech. Partners, LLC, 129 F.4th 1367, 1380 (Fed. Cir. 2025)

“In contrast to the plain and ordinary meaning, Gesture’s claim construction appears to add a temporal limitation to the claims that is simply not there and relies on an argument that claim 7’s ‘reflected light’ implicitly provides an antecedent basis for claim 12’s ‘light source.’ We disagree that claim 7 provides such an antecedent basis.”

The Specification Does Not Support Simultaneity

Example
Reads

Sample
HapMap sa
mately 500
system (Co
was treated
and R01439
according to
ended DNA
Library C
poration: At
the top strai

as AGAGCACACGTC TGAATCCAGTCAC (SEQ ID NO:2), 2) an indexing site, 3) an identifier site having a random 6 base sequence and, 4) a sequence compatible with an Illumina forward sequencing priming site such as TCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO:3), was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to supplier's recommendations. Following 30 minutes of incubation at 25° C., the reaction mixture was diluted with water, 0.8x volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, Mass.) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95° C. then slowly cooling the mixture from 80° C. to 60° C. by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, Mass.) according to the manufacturer's protocols. Following extension, the DNA fragments were col-

Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this

The first file contained the first 8 bases of the indexing reads and is utilized as the library index file for standard library parsing. The other file contains only the random bases and is set aside for further sequence parsing.

Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. FIG. 5 provides the analysis results demonstrating the identification of the duplicate

random reads, the apparent duplicates were found to comprise 39% of all reads. However, the true duplicates, identified through the use of an identifier site, were found to comprise only 26% of all reads. The methods employing the use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

the determi-
the number

tical, their
d occur by
is used in
an be deter-
). With the
ility of two
ese would
atus can be

nds, the use
tion on the
two million

STEP 1

REFERENCE SEQUENCE

A T C C G T A A A T T C G C G

IDENTIFIER

TARGET SEQUENCE

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

The Specification Does Not Support Simultaneity

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

Sample HapMap sample approximately 500 system (Cov was treated and R01439 according to ended DNA Library Corporation: At the top strand as AGAGG NO:2), 2) a random 6 bp an Illumina TCTTTCCGACACGACGCTCTTCCGATCT (SEQ ID NO:3), was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to supplier's recommendations. Following 30 minutes of incubation at 25° C., the reaction mixture was diluted with water, 0.8x volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, Mass.) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95° C. then slowly cooling the mixture from 80° C. to 60° C. by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, Mass.) according to the manufacturer's protocols. Following extension, the DNA fragments were col-

lected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library

reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

If the sequences of two libraries are identical, their

and occur by

is used in

can be deter-

With the

ility of two

ese would

atus can be

ads, the use

tion on the

two million

nd to com-

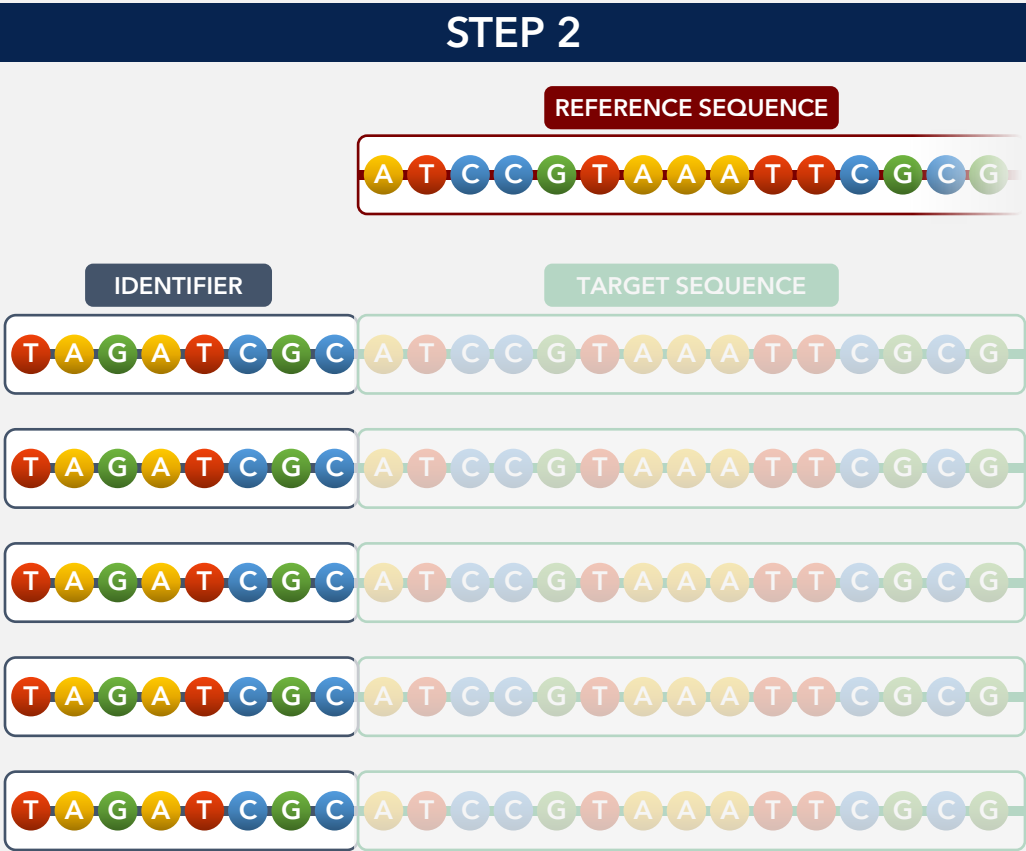
icates, iden-

re found to

employing the

use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random bases that were ligated to them. Where two sequences with the same



The Specification Does Not Support Simultaneity

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

Sample Description: 100 ng of DNA from a human HapMap sample (NA19238) was fragmented to approximately 500 base pair in length by sonication with a Covaris system (Covaris, Inc., Woburn, Mass.). The resulting DNA was treated with end repair enzyme mix NuGEN R01280 and R01439 (NuGEN Technologies, Inc., San Carlos, Calif.)

according to the manufacturer's recommendations. The resulting DNA was ligated to the Ovation Target Enrichment Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to the manufacturer's recommendations. The resulting DNA was ligated to the Ovation Target Enrichment Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to the manufacturer's recommendations. The resulting DNA was ligated to the Ovation Target Enrichment Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to the manufacturer's recommendations.

Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, Calif.) according to supplier's recommendations. Following 30 minutes of incubation at 25° C., the reaction mixture was diluted with water, 0.8x volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, Mass.) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95° C. then slowly cooling the mixture from 80° C. to 60° C. by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, Mass.) according to the manufacturer's protocols. Following extension, the DNA fragments were col-

lected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, Calif.) that also contain the Illumina flow cell sequences (Illumina Inc., San Diego, Calif.) according to supplier's recommendation.

The resulting libraries were quantitated by qPCR using a kit provided by KAPA, diluted to 2 nM and applied to an Illumina MiSeq DNA Sequencer (Illumina Inc., San Diego,

reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would

status can be determined. If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would

were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation

reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. FIG. 5 provides the analysis results demonstrating the identification of the duplicate

STEP 3

REFERENCE SEQUENCE

A T C C G T A A A T T C G C G

IDENTIFIER

TARGET SEQUENCE

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

T A G A T C G C A T C C G T A A A T T C G C G

Example 1 Discloses Identification of “True” Duplicates

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

Sample Description: 100 ng of DNA from a human

lected on Agencourt magnetic beads, washed and eluted according to manufacturer’s recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, Calif.) that

reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in

for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. FIG. 5 provides the analysis results demonstrating the identification of the duplicate reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

mixture was diluted with water, 0.8x volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, Mass.) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated DNA fragments eluted according to manufacturer’s recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95° C. then slowly cooling the mixture from 80° C. to 60° C. by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, Mass.) according to the manufacturer’s protocols. Following extension, the DNA fragments were col-

were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. FIG. 5 provides the analysis results demonstrating the identification of the duplicate

Resolution of apparent duplicate reads by identifier sequence ligation

Libraries were generated from 100ng genomic DNA Covaris fragmented to ~500bp in length. These were enriched with the 344 gene cancer panel, ~12k probes covering 969kb of genomic space and sequenced on the ILMN Miseq. Different numbers of reads were randomly selected for this analysis

Total reads randomly selected	Aligned read	Fold Target Coverage	Total duplicates after identifier analysis (true rate)	Percentage duplicates after identifier analysis (true rate)	Total apparent duplicates with no identifier	Percentage apparent duplicates with no identifier
100K	97,919	2X	1854	1.89	3256	3.33
500K	488,423	7X	39731	8.13	68741	14.07
1M	976,662	13X	145589	14.91	237882	24.36
2M	1,954,673	21X	514848	26.34	764825	39.13
10M	9,771,462	43X	6493577	66.45	7515496	76.91

“Apparent” Duplicates – individual targets within libraries consist of one random end (result of fragmentation) and one common end (initiated with the sequence specific probe). Apparent duplicates are derived from sequencing libraries and using conventional measures of duplicate reads ie, reads were mapped using bowtie and all reads with the same start and end genomic coordinates were counted as duplicates.

“True” duplicates - An identifier sequence is introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates. The analysis method is described in the specification.

Figure 5

“True” duplicates - An identifier sequence is introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates. The analysis method is described in the specification.

Figure 5

The Prosecution History Does Not Support Simultaneity

Attorney Docket No. NUGN-001/01US 301735-2003 PATENT
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of: AMORESE, Douglas, Confirmation No.: 7076
et al.
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

Attorney Docket No.: NUGN-001/03US 34079/516

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT: NuGEN Technologies, Inc. ART UNIT: 1637
SERIAL NUMBER: 16/743,604 CONF. NO.: 3249
FILING DATE: January 15, 2020 EXAMINER: Teresa Strzelecka
TITLE: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A
DUPLICATE SEQUENCING READ

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

FILED ELECTRONICALLY

RESPONSE TO NON-FINAL OFFICE ACTION

This Amendment is in response to the Non-Final Office Action mailed December 8, 2020
from the United States Patent and Trademark Office.

Amendments to the Claims begin on page 2.

Remarks begin on page 4.

Applicant believes that no fees are due in connection with this Amendment. However,
Applicant authorizes any required fees to be charged to Deposit Account No. 500369.

IDT Exhibit 1019.0448

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

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.*

This is obviated in the claimed methods by using the identifier and fragment sequence to identify duplicate reads. Moreover, because the fragment sequence is used in conjunction with the identifier to find duplicates, the identifier can be less complex than the “barcode” in McCloskey. The lower complexity of adaptors comprising fewer nucleotides proves more cost-

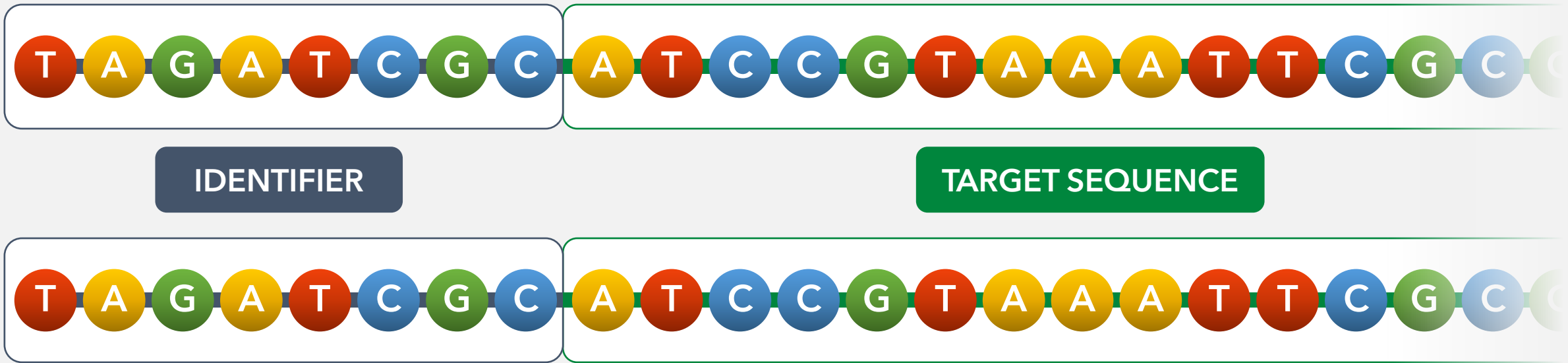
Tecan's Expert Agrees Simultaneity Is Not Required



**Dr. Carlos
Bustamante**
Expert for Tecan

- Q.** And if the comparison is split into two components such that first identifiers are compared and then target sequences are compared, is that still a single step?
- A.** I would say that because the condition is an if and only if, you could start with a string comparison that would begin with a character comparison, right, so it depends on how granular you want to be in defining the recipe by which it's carried out; right? Just like any recipe, you could further break down a step into multiple steps that it comprises. So at the end you need to have unique identifier and – you need to have identical unique identifiers or identical identifiers and identical target sequence in order to identify the sequences as duplicates.
- Q.** And if my workflow compares the identifiers first, then clusters them based on that, then compares target sequences, that is still a single step; correct?
- A.** I would say so, right, and by analogy I could start by comparing the first character and the second character, then the third character and if they are identical you keep going. If they are not identical, you would flag them as not identical and it could be done from left to right or right to left.

Tecan's Expert Agrees Simultaneity Is Not Required



**Dr. Carlos
Bustamante**

- Q.** I'd like to just make sure that's very clear. In your opinion, if the sequence comparison preceded one character at a time, that would still be a single step even though there are end [n] comparisons being made in sequence?
- A.** ... Because string comparison encompasses character comparison, you could go down the representation, right, coded in ASCII and so on; right? So I think it is all part of the single identifier step. You could break it down as far as you would like, it is still that single process by which you're comparing the unique identifier, ensuring that's identical and the target sequencing and ensuring that is identical.

		Tecan	Petitioner
1	Claim Language		
2	Specification		
3	Prosecution History		
4	Dr. Bustamante		

AquaTex Indus., Inc. v. Techniche Sols., 419 F.3d 1374, 1381 (Fed. Cir. 2005)

“In interpreting ‘fiberfill,’ the trial court relied heavily upon statements made during prosecution . . . as a clear disavowal of natural fibers for use as ‘fiberfill batting material.’ The prosecution history, however, is ambiguous and does not directly address the composition of ‘fiberfill.’ Thus, we disagree with the trial court that AquaTex’s arguments foreclose potential compositions of fiberfill, and we decline to give the prosecution history much weight. The representations made during prosecution neither support nor discredit any particular meaning of the term ‘fiberfill.’ They are not a clear disavowal of claim scope.”

			
3	Prosecution History		
4	Dr. Bustamante		

"At Least a Portion" Is Unsupported By the Plain Language

Claim Term	Tecan's Construction
"identifying sequence reads with identical identifier and target sequences as duplicates" '357 Patent, Claim 1	"identifying sequence reads as duplicates ~~through~~ <u>a step</u> that uses both the ~~with-identical~~ identifier and <u>at least a portion of the</u> target sequences ~~as duplicates~~"

Construing the Computer-Aided Limitation in Claim 1

'357 Patent



US011098357B2

(12) **United States Patent**
Amorese et al.

(10) **Patent No.:** US 11,098,357 B2
(45) **Date of Patent:** *Aug. 24, 2021

(54) **COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ**

(71) **Applicant:** NUGEN TECHNOLOGIES, INC., Redwood City, CA (US)

(72) **Inventors:** Douglas A. Amorese, Los Altos, CA (US); Jonathan Seidnick, San Francisco, CA (US); Ben Schroeder, San Mateo, CA (US)

(73) **Assignee:** TECAN GENOMICS, INC., Redwood City, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.
This patent is subject to a terminal disclaimer.

(21) **Appl. No.:** 16/743,604

(22) **Filed:** Jan. 15, 2020

(65) **Prior Publication Data**

US 2020/0239953 A1 Jul. 30, 2020

Related U.S. Application Data

(66) Continuation of application No. 15/406,002, filed on Jan. 13, 2017, now Pat. No. 10,570,448, which is a division of application No. 14/540,917, filed on Nov. 13, 2014, now Pat. No. 9,546,399.

(66) Provisional application No. 61/903,826, filed on Nov. 13, 2013.

(51) **Int. Cl.**

C12Q 1/6874 (2018.01)

C12Q 1/6869 (2018.01)

(52) **U.S. Cl.** C12Q 1/6874 (2013.01); C12Q 1/6869 (2013.01); C12Q 2525/191 (2013.01); C12Q 2525/192 (2013.01)

(58) **Field of Classification Search**

CPC C12Q 1/6869; C12Q 2525/191; C12Q 2525/192

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,362,867 A 12/1982 Paddock
4,458,066 A 7/1984 Caruthers et al.
4,469,863 A 9/1984 Ts'o et al.
4,582,877 A 4/1986 Fairchild et al.
4,876,187 A 10/1989 Duck et al.
4,923,063 A 5/1990 Golias
4,935,557 A 6/1990 Seybold
4,942,124 A 7/1990 Church
4,988,617 A 1/1991 Landegren et al.
4,996,143 A 2/1991 Haller et al.
5,011,769 A 4/1991 Duck et al.
5,034,506 A 7/1991 Summerton et al.

5,035,996 A 7/1991 Hartley

5,043,272 A 8/1991 Hartley

5,082,830 A 1/1992 Parlati et al.

5,090,591 A 2/1992 Long

5,130,238 A 7/1992 Malik et al.

5,169,766 A 12/1992 Schuster et al.

5,171,534 A 12/1992 Smith et al.

5,194,370 A 3/1993 Berninger et al.

5,216,141 A 6/1993 Benner

5,234,409 A 8/1993 Biron et al.

5,235,033 A 8/1993 Summerton et al.

5,242,794 A 9/1993 Whitley et al.

5,384,242 A 1/1995 Oakes

5,386,023 A 1/1995 Sanghvi et al.

5,399,491 A 3/1995 Kacian et al.

5,409,818 A 4/1995 Darvey et al.

5,418,149 A 5/1995 Cellulard et al.

5,422,271 A 6/1995 Chen et al.

5,427,929 A 6/1995 Richards et al.

5,480,784 A 1/1996 Kacian et al.

5,494,810 A 2/1996 Darvey et al.

5,508,169 A 4/1996 Dragan et al.

5,508,178 A 4/1996 Rose et al.

5,510,270 A 4/1996 Foster et al.

5,525,471 A 6/1996 Zeng

5,545,522 A 8/1996 Van Gelder et al.

5,554,516 A 9/1996 Kacian et al.

5,554,517 A 9/1996 Darvey et al.

5,556,752 A 9/1996 Lockhart et al.

5,565,540 A 10/1996 Cheschik et al.

5,573,913 A 11/1996 Rasmussen et al.

5,578,832 A 11/1996 Tridion et al.

5,580,539 A 12/1996 Hampton et al.

5,602,240 A 2/1997 De Moenneker et al.

(Continued)

OTHER PUBLICATIONS

Illumina publication No. 770-2008-011, Multiplexed Sequencing with the Illumina Genome Analyzer System, pp. 1-4 (Year 2008) *

Vaidi et al., Next generation sequencing for molecular diagnosis of neurological diseases, *Acta Neuropathol.*, vol. 126, pp. 275-285 plus supplemental material pp. 1-10 (Year 2012) *

Amos, 2010, DNA pooling in mutation detection with reference to sequence analysis, *Am J Hum Genet* 66:1689-1692.

Belloso, 2014, cnaCapSeq: detecting copy number variation in long-range targeted resequencing data, *Nucleic Acids Res* 42:201-158.

Benson, 2013, Genbank, *Nucl Acids Res* 41:D36-D42.

Bloomquist, 2013, Targeted RNA Sequencing with Competitive Multiplex-PCR, *Amplification Libraries*, PLOS ONE 8(11):e79120.

Bloch, 2013, Comparison of commercially available target enrichment methods for next-generation sequencing, *J Biomolecular Tech* 24:73-86.

(Continued)

Primary Examiner — Teresa E Strezelska

(74) *Attorney, Agent, or Firm* — Brown Rudnick LLP;

Thomas C. Meyers

(57) **ABSTRACT**

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed.

10 Claims, 6 Drawing Sheets

Specification includes a Sequence Listing.

IDT Exhibit 1004.0001

1. A method for detecting duplicate sequencing reads, the method comprising:

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a unique identifier having from about 1 to about 8 nucleotides, an indexing site unique to a subset of the adaptors, and a primer binding site;

amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences; and identifying sequence reads with identical identifier and target sequences as duplicates.

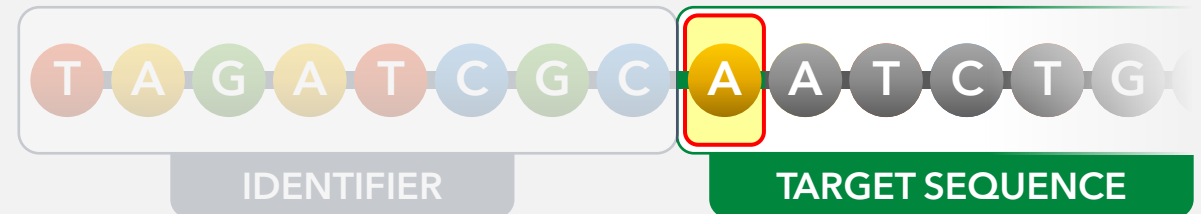
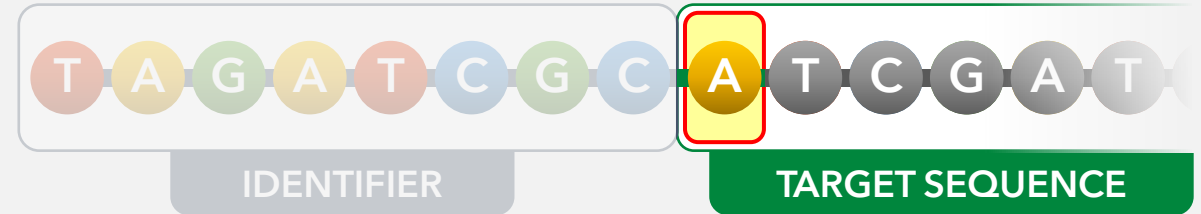
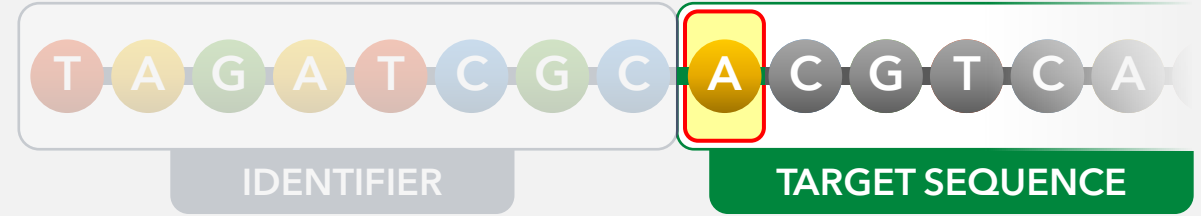
"At Least a Portion" of the Target Sequence Renders Absurd Results



Dr. Carlos Bustamante
Tecan's Expert

Q. So can a single base be a portion of a target sequence?

A. A single base by definition would be a portion of a target sequence but in general, the target sequence would be longer than a single base.



Petitioner's Hearing Plan

1 Claim Construction

2 **Anticipation – Bielas** (EX1013)

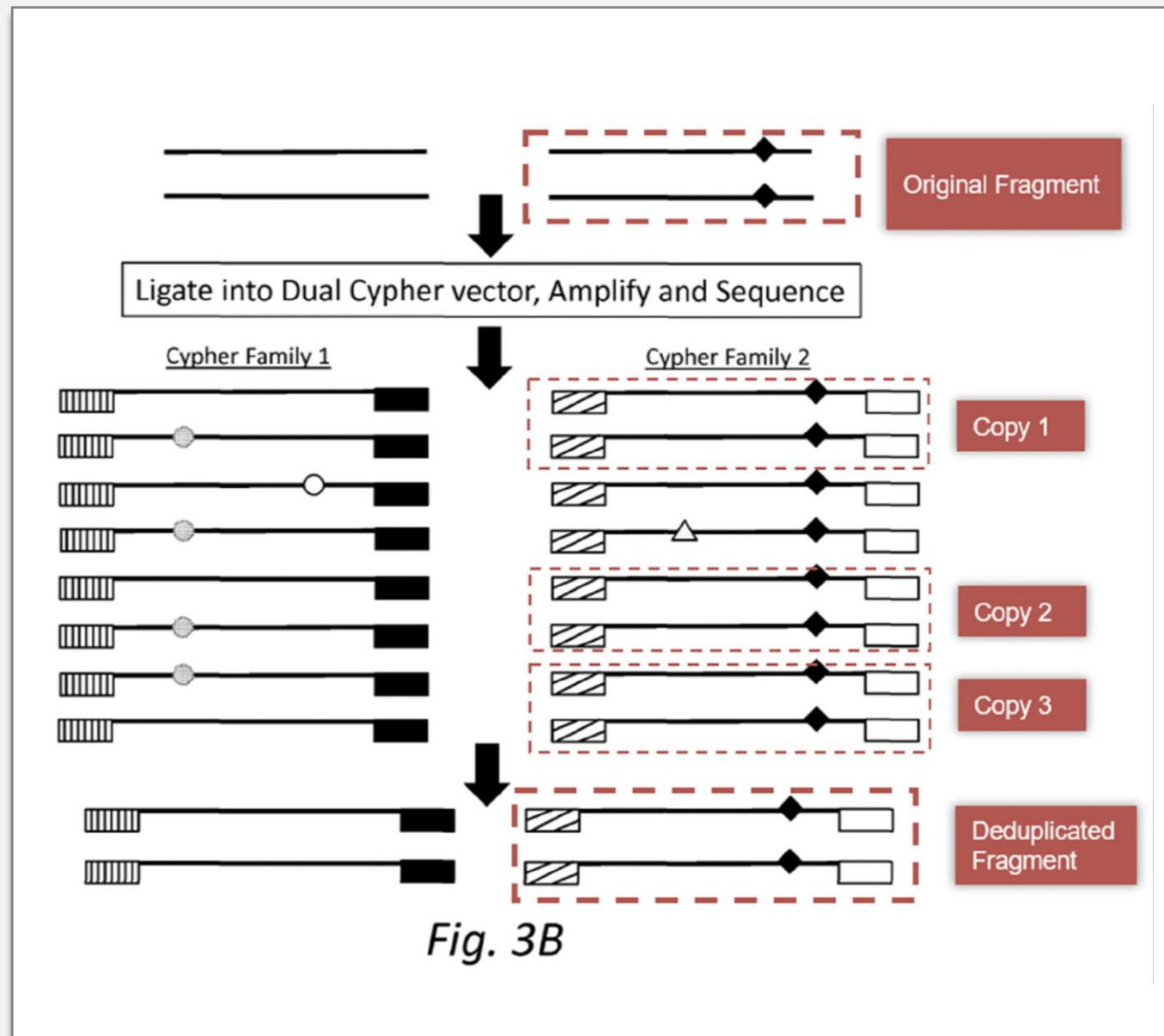
3 Obviousness – Iafrate & Kivioja (EX1011 & 1012)

Bielas's (EX1013) Example 3 Teaches the Method of Claims 1 and 2

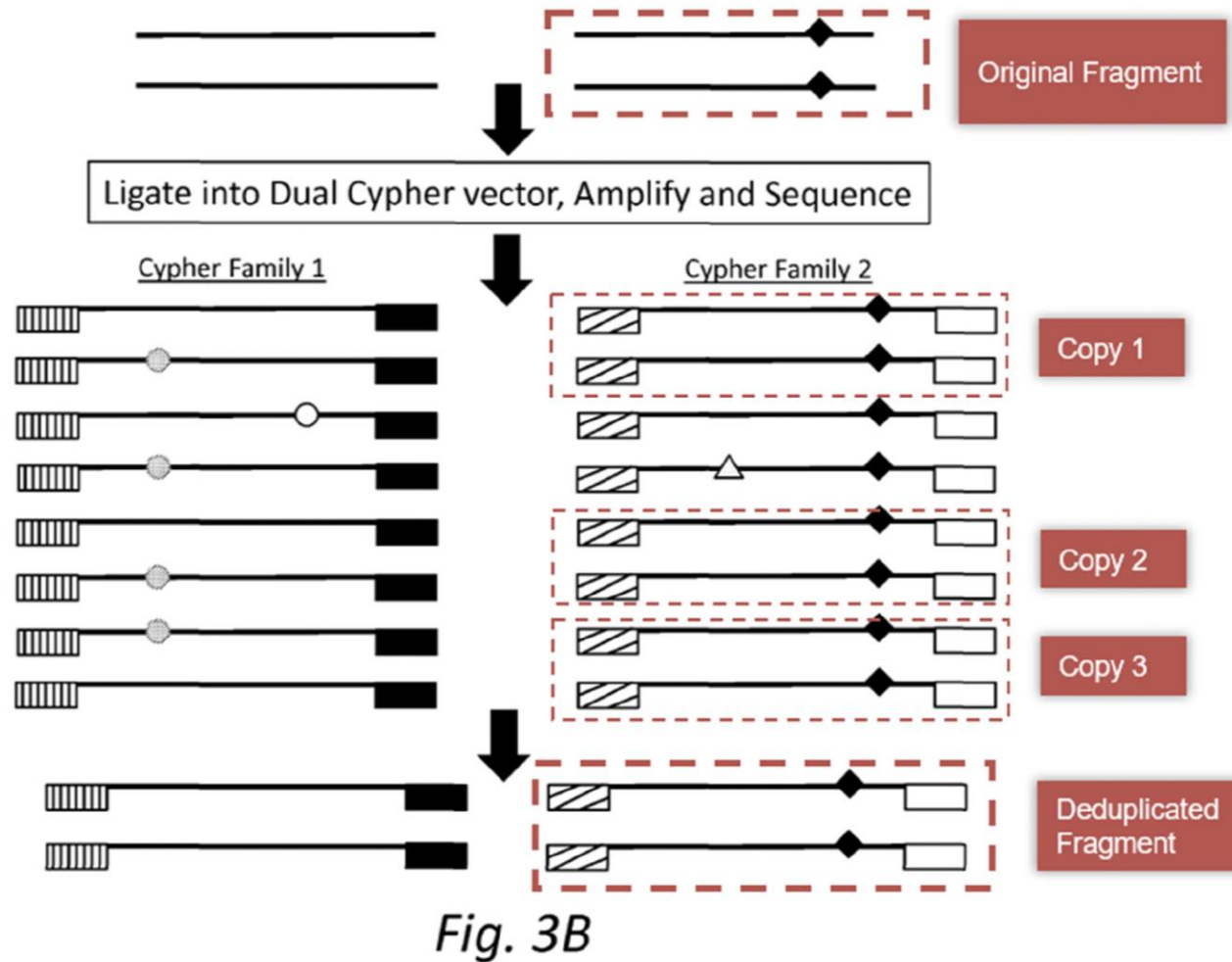
Petitioner's Expert (Dr. Sims)

483. After sequencing, Bielas explains that “all sequencing reads with identical cypher pairs and their reverse complements were grouped into families to create a consensus sequence.” EX1013 (Bielas) at 0031. Bielas diagrammed this process in Figure 3B, which I have reproduced below. EX1013 (Bielas) at 0043.

484. Bielas explains that the “open circle” represents errors introduced during library preparation; the “grey circle and triangle” represent errors introduced during sequencing; and the “black diamond” represents true mutations. EX1013 (Bielas) at 0031–32, 0043. Figure 3B shows that Bielas amplified and sequenced adaptor-ligated fragments and then deduplicated them based on the Cypher and sequence of the nucleic acid fragment. EX1013 (Bielas) at 0031–32.



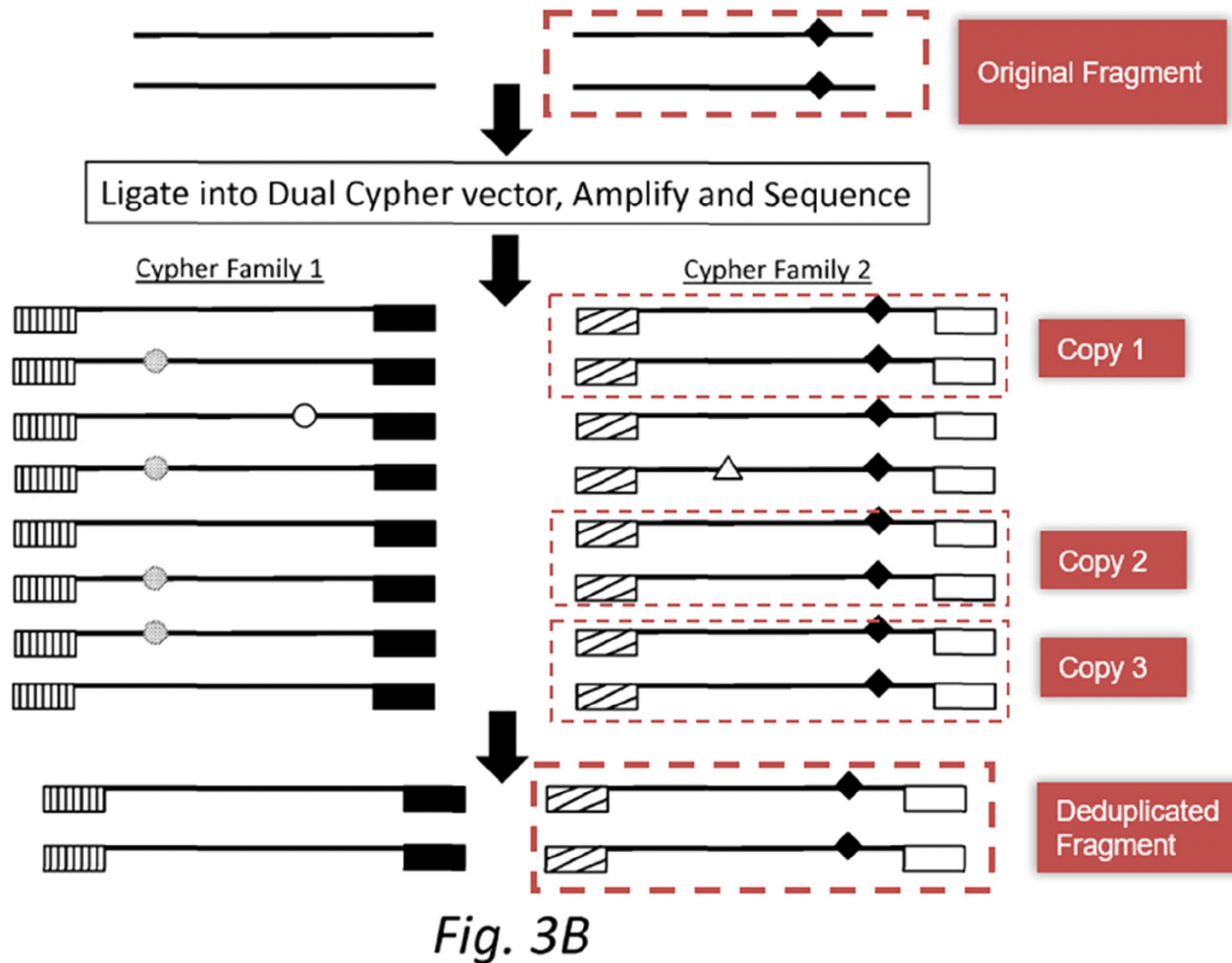
Dr. Bustamante Agrees Bielas Groups Sequences by Identifiers



Tecan's Expert (Dr. Bustamante)

- Q.** We see in cypher family 2 a representation of four different sequence pairs; correct?
- A.** Correct.
- Q.** And these four sequence pairs were grouped into a cypher family because they have the same identifier sequences; correct?
- A.** Or the reverse complement as the case may be, yes.

Dr. Bustamante Agrees Bielas Compares Target Sequences



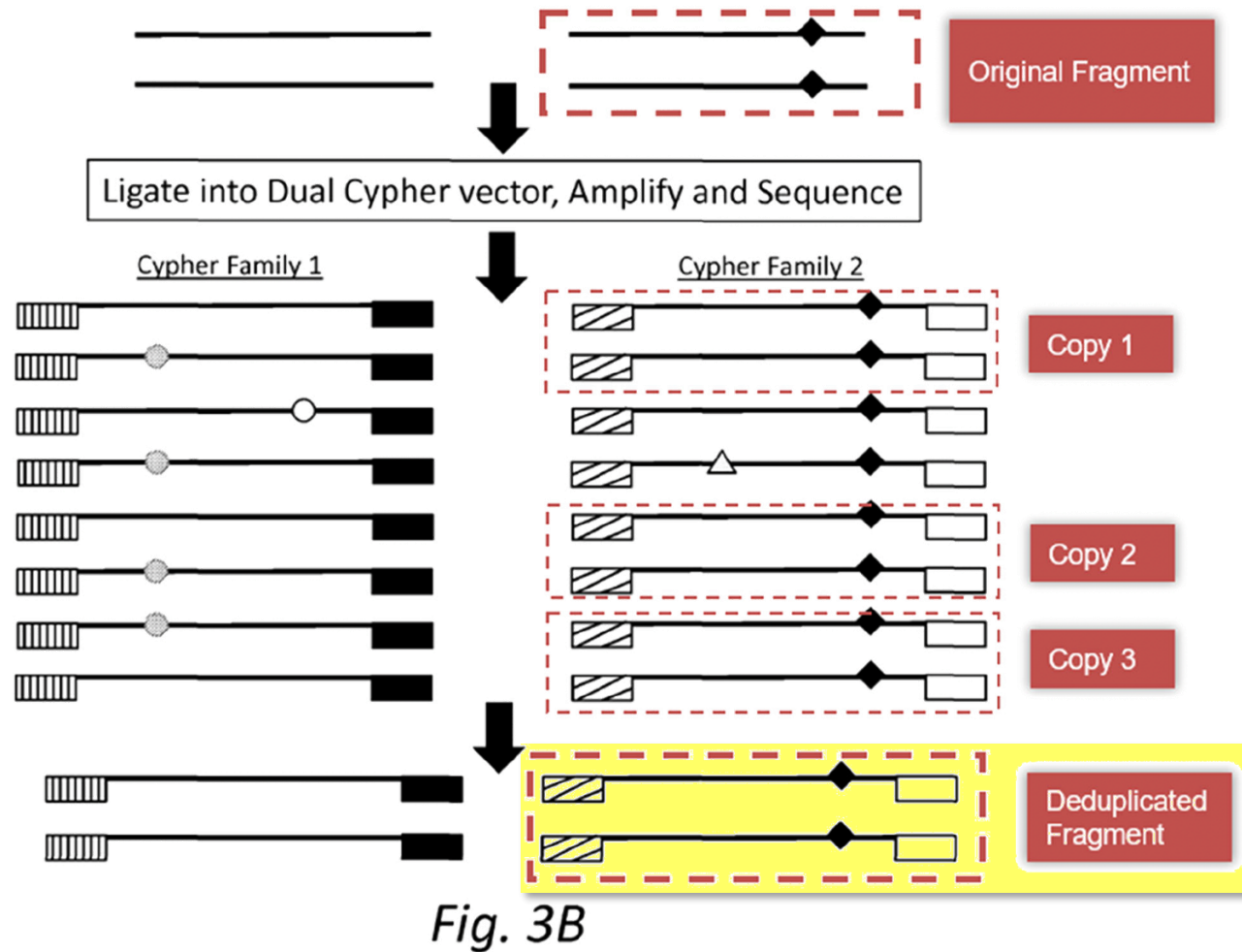
Tecan's Expert (Dr. Bustamante)

- Q.** And it then compares the sequences of the inserts; correct?
- A.** It then processes the sequences of the inserts to generate a consensus.
- Q.** And that process involves comparing the sequences; correct?
- A.** It involves generating a consensus sequence by the method.
- Q.** Of comparing sequences?
- A.** Of comparing positions in the sequences.

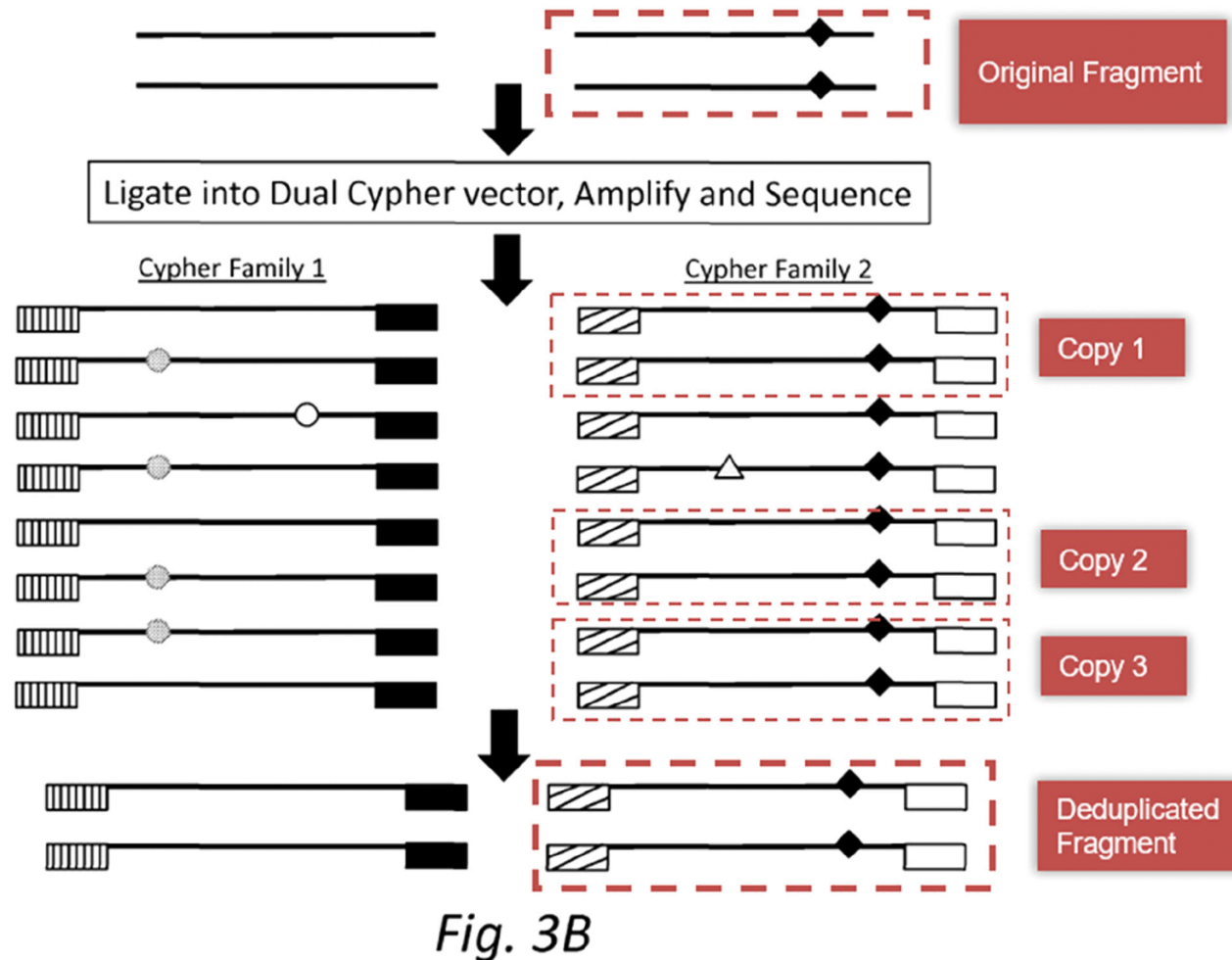
Dr. Bustamante Agrees Bielas Generates a Single Sequence

Tecan's Expert (Dr. Bustamante)

- Q.** The final product of Figure 3B's method is a single pair of sequences; correct?
- A.** A single pair of sequences.
A single pair of consensus sequences, pair of consensus sequences.



Bielas' Data Processing Occurs in a Single "Step"



Tecan's Expert (Dr. Bustamante)

- Q.** I'd like to just make sure that's very clear. In your opinion, if the sequence comparison preceded one character at a time, that would still be a single step even though there are end [n] comparisons being made in sequence?
- A.** It's an if and only - Because string comparison encompasses character comparison, you could go down the representation, right, coded in ASCII and so on; right? So I think it is all part of the single identifier step. You could break it down as far as you would like, it is still that single process by which you're comparing the unique identifier, ensuring that's identical and the target sequencing and ensuring that is identical.

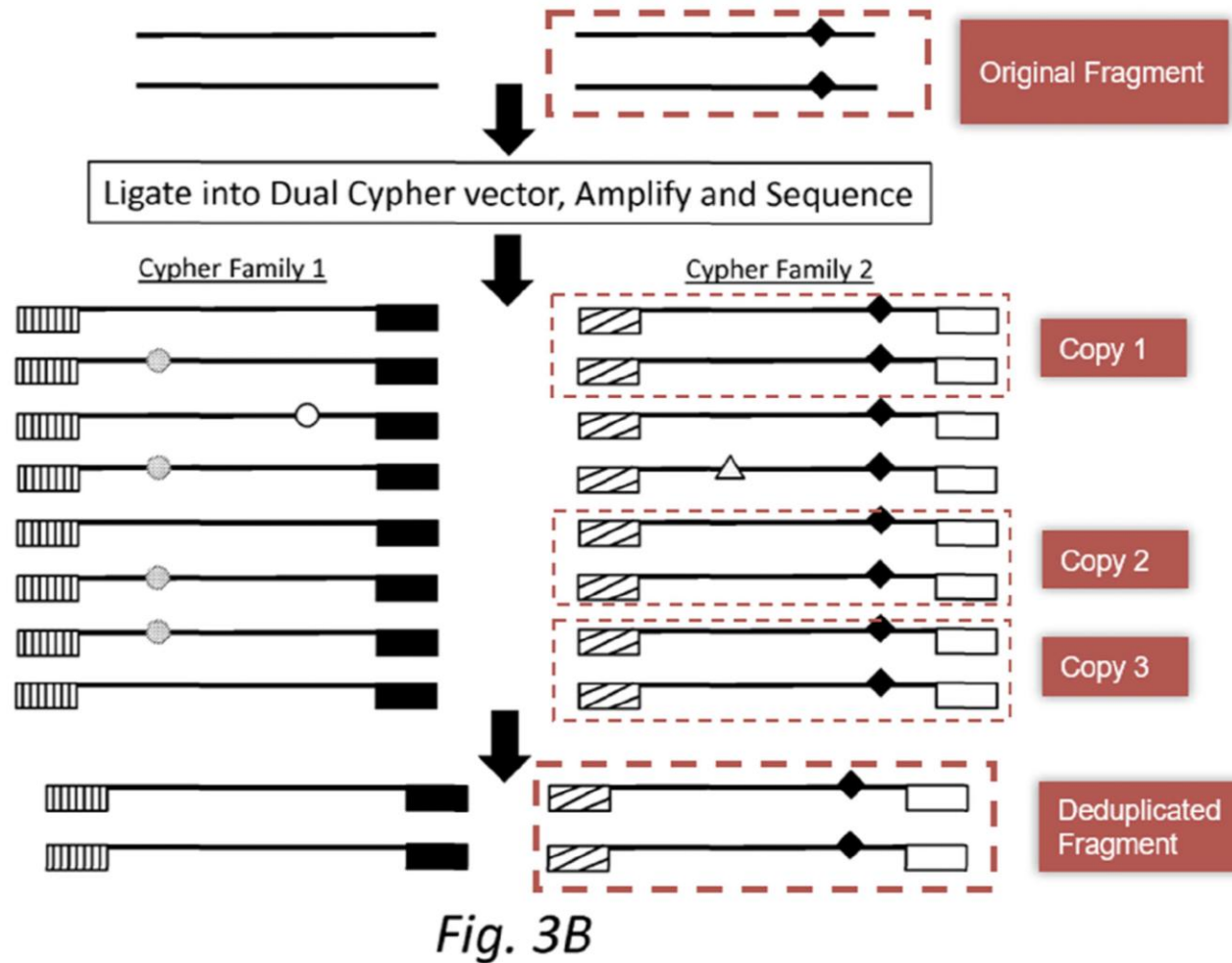
Hewlett-Packard Co. v. Mustek Sys., 340 F.3d 1314, 1324 n.6 (Fed. Cir. 2003)

“The anticipation analysis asks solely whether the prior art reference discloses and enables the claimed invention, and not how the prior art characterizes that disclosure[.]”

Dr. Bustamante Agrees That Generating a Consensus Sequence Can Identify and Remove Duplicates

Tecan's Expert (Dr. Bustamante)

In my opinion, the calculation of a consensus sequence, whatever it happens to be, does not take away from the fact that duplicate reads would be identified using the unique identifier that are identical in target sequences that are identical, unique identifiers that are identical and target sequences that are identical in at least a pair wise fashion. So there are multiple ways that can be accomplished. You start with a pair, define those as identical, add another sequence that is identical to that, that would arrive at the same conclusion regarding duplication that would also have the same consensus sequence. So simply by calculating a consensus sequence you, in my opinion, do not get around the fact that these have unique identifier and unique target sequence in a pair wise fashion and in a multi way fashion.



Tecan Does Not Dispute the Obviousness of an 8-Base Identifier

Patent '357
Claim 9

The method of claim 1, wherein each unique identifier is 8 nucleotides in length.



**Dr. Carlos
Bustamante**

- Q.** So you agree that the person of ordinary skill in the art would have wanted to use the shortest possible molecular barcode that they could in view of the complexity of their sample, given this real estate problem?
- A.** I would say there's a trade-off between the length of the barcode and its ability to uniquely tag the population of molecules and the length of a sequencing read that is allocated to target sequence identification.
- Q.** And that was a recognized trade-off as of 2012; correct?
- A.** You have a fixed amount of real estate and so you have to decide how much is barcode and how much is target sequence.

Petitioner's Hearing Plan

1 Claim Construction

2 Anticipation – Bielas (EX1013)

3 Obviousness – Iafrate & Kivioja (EX1011 & 1012)



US 20130303461A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2013/0303461 A1**
(43) **Pub. Date: Nov. 14, 2013**

(54) **METHODS FOR DETERMINING A NUCLEOTIDE SEQUENCE**

Publication Classification

(71) Applicant: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

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

(72) Inventors: **Anthony John Iafrate**, Newton, MA (US); **Long Phi Le**, Boston, MA (US); **Zongli Zheng**, Boston, MA (US)

(52) **U.S. Cl.**
CPC **C12Q 1/6874** (2013.01)
USPC **514/19.3**; 506/2; 514/318; 514/86;
514/252.14; 514/232.8; 514/183; 514/245;
514/252.02; 514/265.1; 514/414; 514/346;
514/266.22; 514/312; 514/252.18; 514/275;
514/338; 514/333; 514/234.5; 514/172

(73) Assignee: **THE GENERAL HOSPITAL CORPORATION**, Boston, MA (US)

(21) Appl. No.: **13/793,564**

(22) Filed: **Mar. 11, 2013**

(57) **ABSTRACT**

Related U.S. Application Data

(60) Provisional application No. 61/645,364, filed on May 10, 2012, provisional application No. 61/679,302, filed on Aug. 3, 2012.

The technology described herein is directed to methods of determining oligonucleotide sequences, e.g. by enriching target sequences prior to sequencing the sequences.

Inventors: **Anthony John Iafrate**, Newton, MA (US); **Long Phi Le**, Boston, MA (US); **Zongli Zheng**, Boston, MA (US)

BRIEF COMMUNICATIONS

Counting absolute numbers of molecules using unique molecular identifiers

Teemu Kivioja^{1-3,5}, Anna Vähärautio^{1,3,5}, Kasper Karlsson⁴, Martin Bonke¹, Martin Enge³, Sten Linnarsson⁴ & Jussi Taipale³

Counting individual RNA or DNA molecules is difficult because they are hard to copy quantitatively for detection. To overcome this limitation, we applied unique molecular identifiers (UMIs), which make each molecule in a population distinct, to genome-scale human karyotyping and mRNA sequencing in *Drosophila*

sequence. We designate the resulting sequences that can be used to uniquely identify copies derived from each molecule unique molecular identifiers (UMIs; Fig. 1).

As long as the complexity of the library of molecules is maintained, the library can be amplified, normalized or otherwise processed without loss of information about the original molecule count because the number of UMIs in the library acts as a molecular memory of the number of molecules in the starting sample (Supplementary Fig. 1). Upon deep sequencing, each UMI will be observed multiple times, and the number of original DNA molecules can be determined simply by counting each UMI only once. However, long before all UMIs are observed, increasingly precise estimates of the absolute molecule number can be made (Online Methods). This is in contrast to other counting methods, such as direct single-molecule sequencing¹⁻³, which require that all counted molecules are observed directly. In addition, many

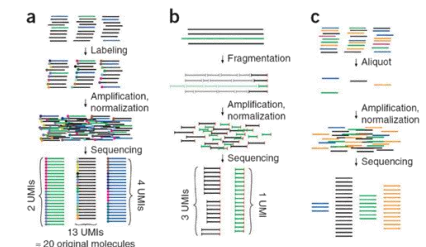
Teemu Kivioja^{1-3,5}, Anna Vähärautio^{1,3,5}, Kasper Karlsson⁴, Martin Bonke¹, Martin Enge³, Sten Linnarsson⁴ & Jussi Taipale³



is small enough to contain only distinct molecules (Fig. 1a-c). Any combination of these manipulations can be used to generate a library in which each molecule has a distinct identifying

(Supplementary Fig. 1). To assess whether it can be used to improve experimental measurements, we applied UMI counting to digital karyotyping and mRNA sequencing (mRNA-seq).

Figure 1 | UMIs can be generated by adding oligonucleotide labels, fragmenting, taking a small enough aliquot or a combination thereof. (a) Three different DNA species (green, blue and black lines) are labeled with a collection of random labels (colored filled circles). Two green molecules are originally present (top), corresponding to two different UMIs (red, blue) among the sequenced molecules (green; bottom). Information about the original number of molecules (top) is preserved in the number of different UMIs detected by sequencing a sample of the amplified and normalized library (bottom). Even if some UMIs are not observed, the original number of molecules can be estimated using count statistics. (b) The original molecule is randomly fragmented, and a short unique sequence from the resulting fragments constitutes each UMI; here only the fragment adjacent to the poly(A) sequence (red vertical bars) is amplified. (c) An aliquot is taken from a sample that has many identical molecules such that on average, less than one copy of each molecule remains.



¹Genome-Scale Biology Program, Institute of Biomedicine, University of Helsinki, Helsinki, Finland. ²Department of Computer Science, University of Helsinki, Helsinki, Finland. ³Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. ⁴Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. ⁵These authors contributed equally to this work. Correspondence should be addressed to J.T. (jussi.taipale@ki.se) or S.L. (sten.linnarsson@ki.se).

RECEIVED 10 MARCH; ACCEPTED 27 SEPTEMBER; PUBLISHED ONLINE 20 NOVEMBER 2011; DOI:10.1038/NMETH.1778

The POSA Would Have Been Motivated to Use lafrate's NGS Library Preparation Method

lafrate

	
US 20130303461A1	
(19) United States	(10) Pub. No.: US 2013/0303461 A1
(12) Patent Application Publication	(43) Pub. Date: Nov. 14, 2013
lafrate et al.	
(54) METHODS FOR DETERMINING A NUCLEOTIDE SEQUENCE	Publication Classification
(71) Applicant: THE GENERAL HOSPITAL CORPORATION, Boston, MA (US)	(51) Int. Cl. C12Q 1/68 (2006.01)
(72) Inventors: Anthony John lafrate, Newton, MA (US); Long Phi Le, Boston, MA (US); Zongli Zheng, Boston, MA (US)	(52) U.S. Cl. CPC: C12Q 1/6874 (2013.01); USPC: 514/193, 506.2; 514/318; 514/86; 514/252.14; 514/232.8; 514/183; 514/245; 514/252.02; 514/265.1; 514/414; 514/346; 514/266.22; 514/312; 514/252.18; 514/275; 514/338; 514/333; 514/234.5; 514/172
(73) Assignee: THE GENERAL HOSPITAL CORPORATION, Boston, MA (US)	
(21) Appl. No.: 13/793,564	
(22) Filed: Mar. 13, 2013	(57) ABSTRACT
Related U.S. Application Data	
(60) Provisional application No. 61/645,364, filed on May 10, 2012; provisional application No. 61/679,302, filed on Aug. 3, 2012.	
The technology described herein is directed to methods of determining oligonucleotide sequences, e.g. by enriching target sequences prior to sequencing the sequences.	

IDT Exhibit 1011.0001

The technology described herein is directed to methods of determining oligonucleotide sequences, e.g. by enriching target sequences prior to sequencing the sequences.

BACKGROUND

[0005] Target enrichment prior to next-generation sequencing is more cost-effective than whole genome, whole exome, and whole transcriptome sequencing and therefore more practical for broad implementation; both for research discovery and clinical applications. For example, high coverage

The POSA Would Have Been Motivated to Modify Iafrate to Improve Its Accuracy



Dr. Peter Sims
Expert for IDT

225. Iafrate teaches that its method of selective enrichment can be used to elucidate information encoded adjacent to a known sequence of interest. EX1011 (Iafrate) at 0012, 0028–30. For example, Iafrate used the method to detect “previously unknown fusion partners” adjacent to the ROS1 gene, which is implicated in lung cancer. EX1011 (Iafrate) at 0029–30 (¶¶233, 241–47). The capability of Iafrate’s method to selectively enrich unknown sequences adjacent to a known gene makes the method particularly suited for cancer research and diagnostics. EX1011 (Iafrate) at 0023–25, 0029. Because Iafrate’s method could be applied to molecular diagnostics and research for cancer, the POSA would have been motivate to make the method as accurate as possible.

The POSA Would Have Been Motivated to Modify Iafrate to Improve Its Accuracy



Dr. Peter Sims
Expert for IDT

229. But Iafrate's method includes several rounds of PCR, which can cause problems for accurately detecting mutations or accurately counting nucleic acid fragments in a sample. In one embodiment, Iafrate calls for "14 cycles of PCR amplification." EX1011 (Iafrate) at 0014. In other embodiments, Iafrate states that "each amplification step can comprise a set of cycles of PCR amplification regimen from 5 cycles to 20 cycles in length." EX1011 (Iafrate) at 00013, 0019, 0027. Because Iafrate's method requires two amplification steps, EX1011 (Iafrate) at 0051, Iafrate requires as many as 40 cycles of PCR.

The POSA Would Have Been Motivated to Modify Iafrate to Improve Its Accuracy



Dr. Peter Sims
Expert for IDT

233. Because the principal application of Iafrate's method is identifying changes in genetic sequences that could help with cancer diagnosis and treatment, the POSA would have wanted to ensure sequencing was as accurate as possible, both in terms of the ability to correctly count original fragments and differentiate them from duplicates, and in terms of the accuracy of sequencing of each individual fragment (accounting for PCR and sequencing error). While true for all of Iafrate's applications, in particular if the POSA applied Iafrate's method to detect mutations in one or a handful of bases, accuracy would have been paramount. Thus, due to the number of cycles of PCR in Iafrate's method, and the associated potential for PCR-introduced error, as well as error introduced by SBS itself, the POSA would have been motivated to modify Iafrate's method to enhance its accuracy.

Iafrate's Adaptor Includes Everything Except Claim 1's "Unique Identifier"

'357 Patent, Claim 1

ligating an adaptor to each of a plurality of nucleic acid fragments, wherein each adaptor comprises a **unique identifier** having from about 1 to about 8 nucleotides, an **indexing site** unique to a subset of the adaptors, and a **primer binding site**;

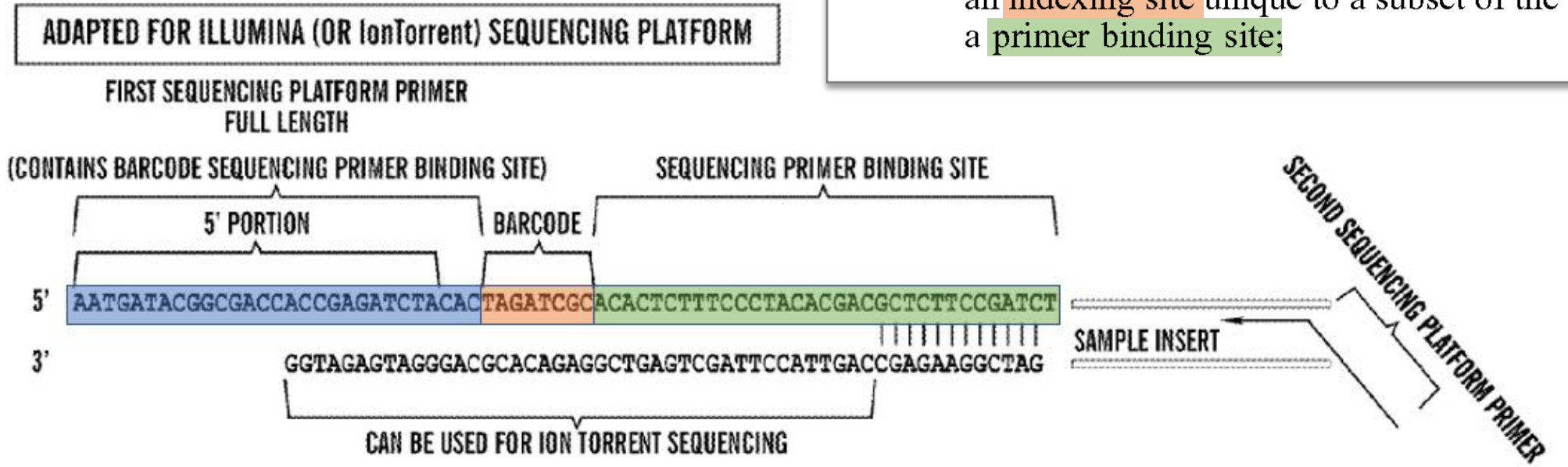


FIG. 9

Kivioja Teaches Methods to Improve Accuracy

Kivioja

BRIEF COMMUNICATIONS

Counting absolute numbers of molecules using unique molecular identifiers

Teemu Kivioja^{1,3,5}, Anna Vähärautio^{1,3,5}, Kasper Karlsson⁴, Martin Bonke¹, Martin Eng³, Sten Linnarsson¹ & Jussi Taipale³

Counting individual RNA or DNA molecules is difficult because they are hard to copy quantitatively for detection. To overcome this limitation, we applied unique molecular identifiers (UMIs), which make each molecule in a population distinct, to genome-scale human karyotyping and mRNA sequencing in *Drosophila melanogaster*. Use of this method can improve accuracy of almost any next-generation sequencing method, including chromatin immunoprecipitation–sequencing, genome assembly, diagnostics and manufacturing–process control and monitoring.

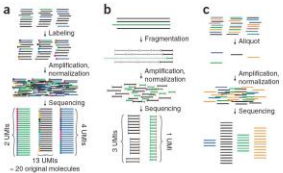
Determining the relative abundance of two different molecular species or the absolute number of molecules in a single sample is challenging. We describe an absolute counting method that can use amplification but does not require detecting each original molecule or keeping track of the number of copies made. In this method, each molecule in a population is first made unique. This can be accomplished by adding a random DNA sequence label, by fragmenting or by taking an aliquot of a complex mixture that is small enough to contain only distinct molecules (Fig. 1a–c). Any combination of these manipulations can be used to generate a library in which each molecule has a distinct identifying

sequence. We designate the resulting sequences that can be used to uniquely identify copies derived from each molecule unique molecular identifiers (UMIs; Fig. 1).

As long as the complexity of the library of molecules is maintained, the library can be amplified, normalized or otherwise processed without loss of information about the original molecule count because the number of UMIs in the library acts as a molecular memory of the number of molecules in the starting sample (Supplementary Fig. 1). Upon deep sequencing, each UMI will be observed multiple times, and the number of original DNA molecules can be determined simply by counting each UMI only once. However, long before all UMIs are observed, increasingly precise estimates of the absolute molecule number can be made (Online Methods). This is in contrast to other counting methods, such as direct single-molecule sequencing^{1–3}, which require that all counted molecules are observed directly. In addition, many existing digital molecule counting methods such as digital PCR⁴, digital microarray profiling⁵ and direct single-molecule sequencing^{1–3} cannot be effectively multiplexed and are thus generally only applicable to measuring one or few molecular species from many samples, or many species from a single sample. Counting methods that introduce random tags to make molecules unique before amplification have been suggested^{5,6} and applied to analysis of RNA–protein interactions^{7,8}. In addition, three recent publications applied such labeling methods to the analysis of selected target genes by using either microarrays⁹ or sequencing^{10,11}. Here we apply the idea more generally and show that it can be used for absolute quantification.

The UMI method is very effective on simulated data (Supplementary Fig. 1). To assess whether it can be used to improve experimental measurements, we applied UMI counting to digital karyotyping and mRNA sequencing (mRNA-seq).

Figure 1 | UMIs can be generated by adding oligonucleotide labels, fragmenting, taking a small enough aliquot or a combination thereof. (a) Three different DNA species (green, blue and black lines) are labeled with a collection of random labels (colored filled circles). Two green molecules are originally present (top), corresponding to two different UMIs (red, blue) among the sequenced molecules (green; bottom). Information about the original number of molecules (top) is preserved in the number of different UMIs detected by sequencing a sample of the amplified and normalized library (bottom). Even if some UMIs are not observed, the original number of molecules can be estimated using count statistics. (b) The original molecule is randomly fragmented, and a short unique sequence from the resulting fragments constitutes each UMI; here only the fragment adjacent to the poly(A) sequence (red vertical bars) is amplified. (c) An aliquot is taken from a sample that has many identical molecules such that on average, less than one copy of each molecule remains.



¹Genome Scale Biology Program, Institute of Biomedicine, University of Helsinki, Helsinki, Finland. ²Department of Computer Science, University of Helsinki, Helsinki, Finland. ³Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. ⁴Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. ⁵These authors contributed equally to this work. Correspondence should be addressed to T.T. (teemu.taipale@ki.se) or S.L. (sten.linnarsson@ki.se).

RECEIVED 10 MARCH; ACCEPTED 27 SEPTEMBER; PUBLISHED ONLINE 20 NOVEMBER 2011; DOI:10.1038/NMETH.1778

72 | VOL.9 NO.1 | JANUARY 2012 | NATURE METHODS
IDT Exhibit 1012.0001

melanogaster. Use of this method can improve accuracy of almost any next-generation sequencing method, including chromatin immunoprecipitation–sequencing, genome assembly, diagnostics and manufacturing–process control and monitoring.

Petitioner's Motivation to Combine Argument Stands Unrebutted



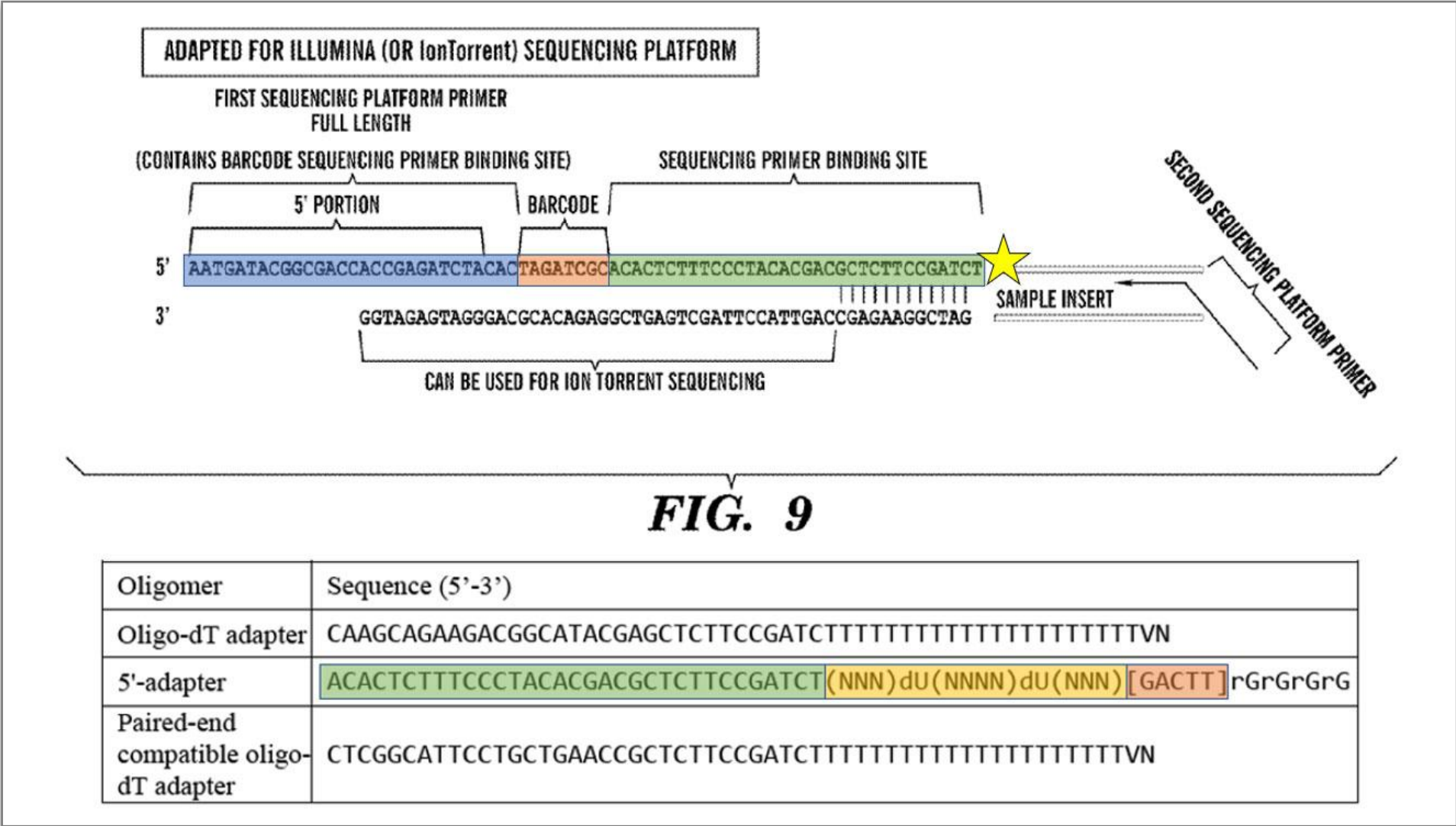
**Dr. Carlos
Bustamante**
Expert for Tecan

- Q.** So just to make sure the record is clear, in your declaration you do not address the motivation or not to combine lafrate and Kivioja; correct?
- A.** I do not offer a specific opinion to that in the declaration.

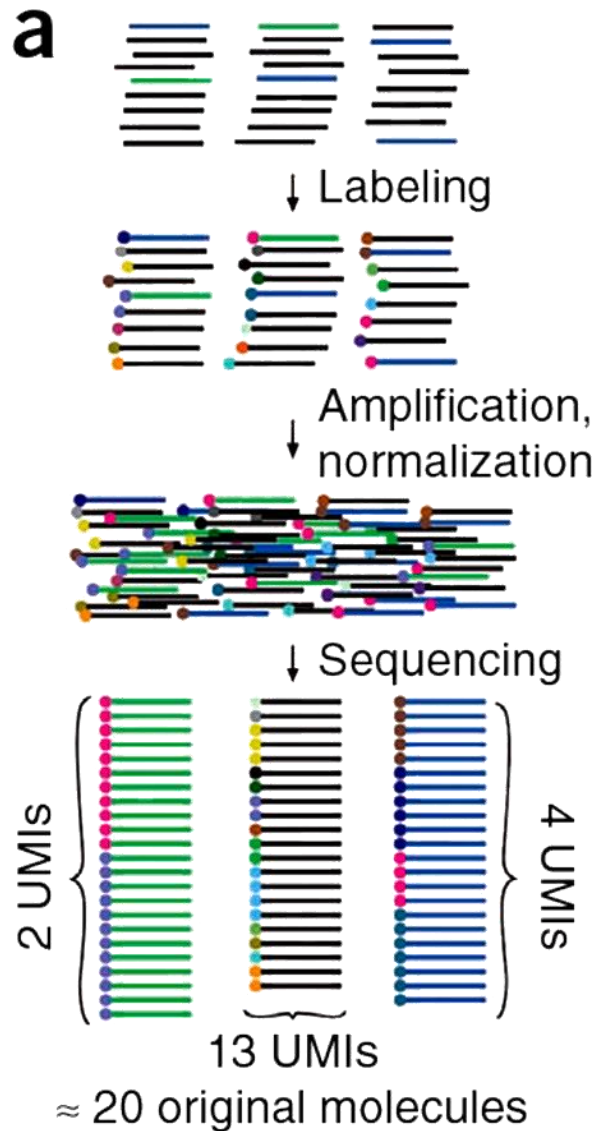
Dr. Sims Explained How the POSA Could Combine Iafrate's Adaptor and Kivioja's Adaptor



Dr. Peter Sims
Expert for IDT

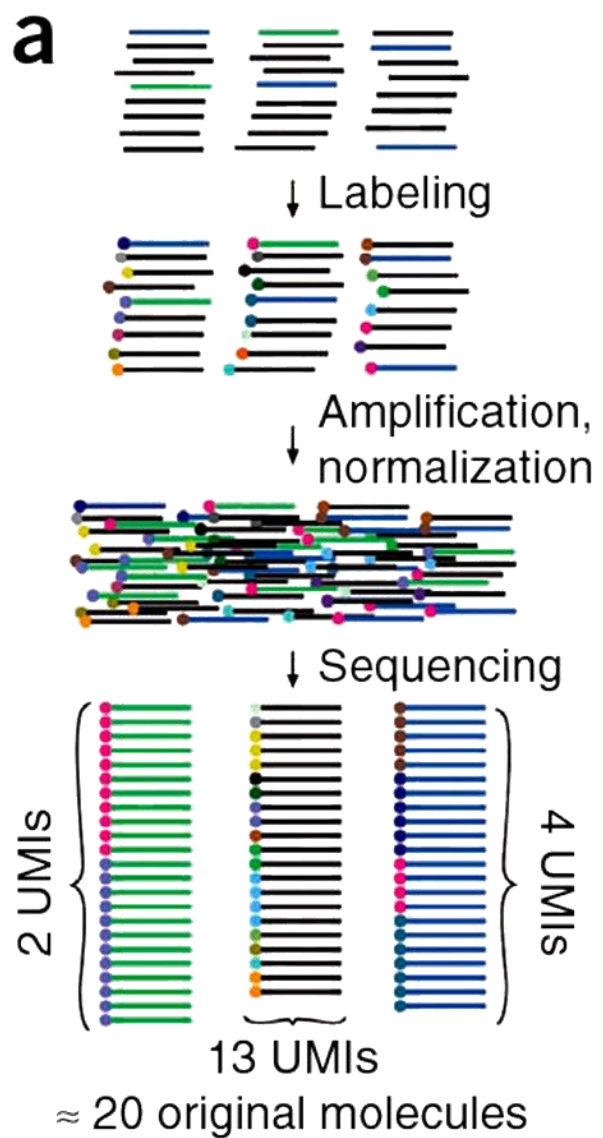


Dr. Bustamante Agrees That Kivioja Identifies Duplicates Using Unique Identifiers



Tecan's Expert (Dr. Bustamante)

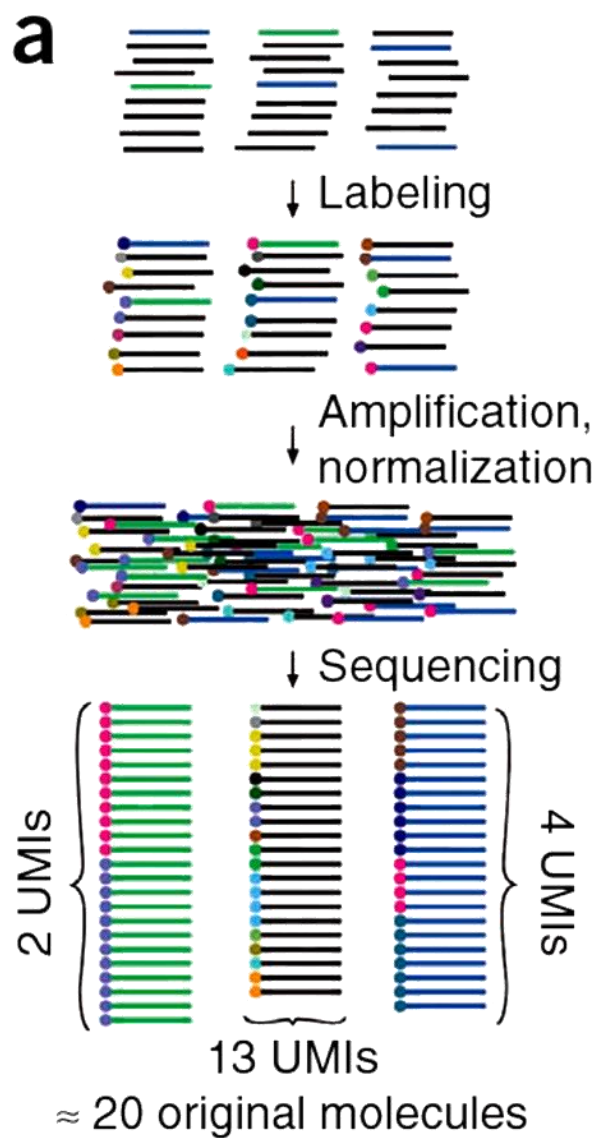
- Q.** So we're clear, Kivioja teaches a method of identifying duplicate sequences using the molecular barcode and the gene and position information; correct?
- A.** That is part of the Kivioja method.



Tecan's Expert (Dr. Bustamante)

- Q.** And the other piece of information that the authors in Kivioja use to identify duplicates is the genomic alignment; correct?
- A.** That is correct.

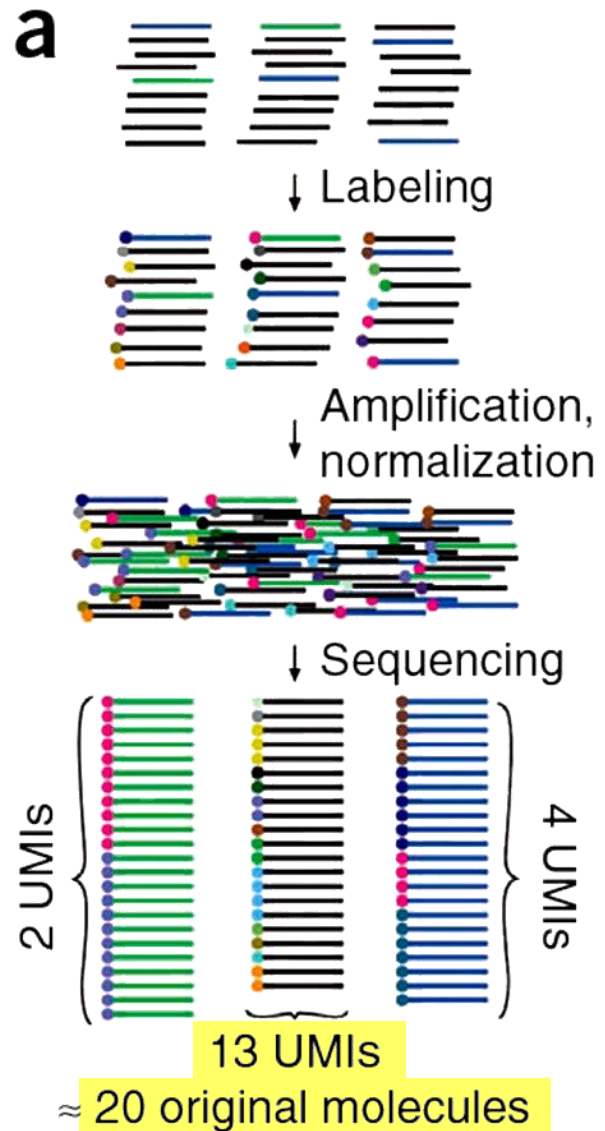
Dr. Bustamante Agrees That Genomic Coordinates Are Obtained Using Target Sequences



Tecan's Expert (Dr. Bustamante)

When aligning sequencing reads against a reference genome (a process termed “mapping”), the alignment tools compare the nucleotide sequence of each sequencing read to the nucleotide sequences contained in the reference genome and identify the location or locations within the reference genome where the sequencing read best matches.

Dr. Bustamante Agrees That Kivioja "Deals With" Its Duplicates



Tecan's Expert (Dr. Bustamante)

- Q.** And counting unique molecules involves figuring out which molecules are duplicates and counting them; correct?
- A.** It involves dealing with duplicate reads one way or another.

CERTIFICATE OF SERVICE

The undersigned hereby certifies that on January 20, 2026, I caused a true and correct copy of the foregoing Petitioner's Demonstrative Exhibits to be served on counsel for Patent Owner via email at:

Timothy J. Beavers	tbeavers@goodwinlaw.com
Linnea P. Cipriano	lcipriano@goodwinlaw.com
John P. Padro	jpadro@goodwinlaw.com
Scott T. Weingaertner	sweingaertner@goodwinlaw.com
	dg-goodwintecan@goodwinlaw.com

Dated: January 20, 2026

Respectfully submitted,

/Thomas S. Fletcher/
Thomas S. Fletcher (Reg. No. 72,383)
WILLIAMS & CONNOLLY LLP
680 Maine Avenue SW
Washington, DC 20024
tfletcher@wc.com

*Lead Counsel for Petitioner
Integrated DNA Technologies, Inc.*