UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

INTEGRATED DNA TECHNOLOGIES, INC.,
Petitioner,

v.

TECAN GENOMICS, INC.,
Patent Owner.

IPR2024-01502
IPR2024-01504
IPR2024-01506

U.S. Patents 9,546,399, 11,098,357,
and 11,725,241

DECLARATION OF PETER A. SIMS, PH.D.

TABLE OF CONTENTS

I. BACKGROUND AND QUALIFICATIONS 1

II. MATERIALS CONSIDERED 8

III. LEGAL STANDARDS 15

A. Burden of Proof 15

B. Anticipation And Obviousness Are Evaluated From The Perspective Of The POSA. 15

C. Anticipation 15

D. Obviousness 16

IV. TECHNICAL BACKGROUND 18

A. DNA 18

B. RNA 24

C. DNA Synthesis 25

D. Genomic DNA 28

E. Ligation and Adaptors 31

F. DNA sequencing 33

G. RNA Sequencing 34

H. "First Generation" Sanger Sequencing 35

I. Polymerase Chain Reaction ("PCR") 38

J. "Anchored PCR" and "Genome Walking" 40

K. "Next Generation Sequencing" 45

1. NGS Platforms 51

a. Roche/454 Pyrosequencing 52

b. Solexa/Illumina Reversible Terminator Sequencing 54

c. Life Technologies/Ion Torrent Semiconductor Sequencing.....59

2. Library Preparation for NGS 60

a. Barcodes for Sample Indexing 61

b. Molecular Tags 66

1) Hug & Schuler 2003.....69

2) Miner 2004 71

3) McCloskey 2007 72

4) Hiatt 2010 75

5) König 2010 78

6) Kinde 2011 80

7) Casbon 2011 83

8) Fu 2011 84

9) Jabara 2011 87

10) Kivioja 2012 (online 2011) 89

11) Tollervey 2011 94

12) Shiroguchi 2012 95

13) Schmitt 2012.....97

14) Later Publications.....100

3. Single-Read and Paired-End Sequencing 101

L. The MGH/Archer "Anchored Multiplex PCR" Library Preparation Method 103

V. THE PERSON OF ORDINARY SKILL IN THE ART 108

VI. THE '399, '357, AND '241 PATENTS 109

A. The Priority Date For The '399, '357, and '241 Patents 112

B. Claim Terms Defined In The '399, '357, and '241 Patents. 112

C. The Challenged Claims Of The '399, '357, And '241 Patents Would Have Been Obvious Over Iafrate And Kivioja. 115

1. Iafrate Is Prior Art. 115

2. Kivioja Is Prior Art. 115

3. The POSA Would Have Been Motivated To Use Iafrate's Method and Adapter for NGS Library Preparation. 116

4. The POSA Would Have Been Motivated to Modify Iafrate's Method To Incorporate Molecular Tags And Use UMIs As Taught By Kivioja. 126

5. The POSA Would Have Had a Reasonable Expectation of Success in Combining Iafrate and Kivioja to Arrive At The Claimed Method. 139

6. Claims 1–3, 5–11, and 14–17 Of The '399 Patent Would Have Been Obvious Over Iafrate and Kivioja. 141

a. Claim 1 141

1) Claim 1: "A method for detecting a duplicate sequence read from a population of sample sequencing reads" 143

2) Claim 1: Step a) 145

3) Claim 1: Steps b) and c) 152

4) Claim 1: Step d) 152

5) Claim 1: "wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides".....155

6) Claim 1: "wherein the sequence of the identifier site is variable in sequence content in the plurality of adapters".....156

7) Claim 1 Would Have Been Obvious Over Iafrate and Kivioja.....157

b. Claim 2: "The method of claim 1, wherein the method further comprises removing from the population of sequencing reads the sequencing read with a duplicate identifier site and target sequence." 157

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the indexing site." 158

d. Claim 5: "The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample." 161

e. Claim 6: "The method of claim 5, wherein the nucleic acid fragments from each sample have the same indexing site and each sample has a unique indexing site." 162

f. Claim 7: "The method of claim 6, wherein the sequencing reads are separated based on the indexing site." 163

g. Claim 8: "The method of claim 7, wherein the separation of sequencing reads is performed prior to step d)." 164

h. Claim 9: "The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments." 165

i. Claim 10: "The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."....166

j. Claim 11: "The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length."....167

k. Claim 14: "The method of claim 1, wherein the identifier site is 8 nucleotides in length."....168

l. Claim 15: "The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site."....173

m. Claim 16: "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site."....175

n. Claim 17: "The method of claim 6, wherein the method further comprises separating fragments or sequencing reads based on indexing site."....176

7. Claims 1-4, 6-7, And 9-10 Of The '357 Patent Would Have Been Obvious Over Iafrate And Kivioja....177

a. Claim 1....179

1) "A method for detecting duplicate sequencing reads, the method comprising:"....180

2) "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"....181

3) "amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences"....183

4) "and identifying sequence reads with identical identifier and target sequences as duplicates." 184

b. Claim 2: "The method of claim 1, further comprising removing a duplicate read from the sequence reads." 185

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence." 186

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence." 190

e. Claim 6: "The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments." 192

f. Claim 7: "The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site." 193

g. Claim 9: "The method of claim 1, wherein each unique identifier is 8 nucleotides in length." 194

h. Claim 10: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic acid fragment." 194

8. Claims 1-16 Of The '241 Patent Were Anticipated By Kivioja. 196

a. Claim 1 196

1) "A method for detecting a duplicate sequence read from a population of sample sequencing reads" 197

2) "obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor".....198

3) "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment".....203

4) "sequencing the amplicons to generate sequence reads that include identifier and target sequence, and".....204

5) "identifying sequence reads with identical identifier and target sequences as duplicates." 205

b. Claim 2: "The method of claim 1, further comprising removing a duplicate read from the sequence of reads." 207

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence." 207

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence." 210

e. Claim 5: "The method of claim 1, wherein the amplicons include amplified fragments of cDNA." 211

f. Claim 6: "The method of claim 1, wherein the amplicons include amplified fragments of genomic DNA." 212

g. Claim 7: "The method of claim 1, wherein each adaptor includes a primer binding site." 213

h. Claim 8: "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site." 214

i. Claim 9: "The method of claim 1, wherein teach unique identifier is at least 6 nucleotides in length." 214

j. Claim 10: "The method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction."....215

k. Claim 11: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment."....216

l. Claim 12: "The method of claim 1, further comprising determining the genomic start position for each sequencing read."....218

m. Claim 13: "The method of claim 1, wherein each adaptor further comprises an indexing site."....219

n. Claim 14: "The method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length."....219

o. Claim 15: "The method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons."....220

p. Claim 16: "The method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index."....221

D. The Challenged Claims Of The ’399, ’357, And ’241 Patents Are Anticipated And/Or Obvious Over Bielas. 222

1. Bielas Disclosed Methods With All Of The Elements Claimed In The ’399, ’357, And ’241 Patents. 222

a. Bielas Example 1 225

b. Bielas Example 3 231

2. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The ’399 Patent 244

a. Bielas Anticipates Claim 1. 244

1) “A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:” 244

2) “a) ligating an adapter to a 5′ end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples,” 245

i. “wherein the adapter comprises: (i) an indexing primer binding site;” 247

ii. “(ii) an indexing site;” 247

iii. “(iii) an identifier site consisting of between 1 and 8 nucleotides; and 249

iv. “(iv) a target sequence primer binding site;” 251

3) “(b) amplifying the adapter-nucleic acid fragment ligated products;” 252

4) “(c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and” 252

5) “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;”.....253

6) “wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides;”.....254

7) “and wherein the sequence of the identifier site is variable in sequence content in a plurality of adaptors.”.....255

b. To The Extent Bielas Does Not Anticipate Claim 1, Claim 1 Would Have Been Obvious Based On Bielas. 256

c. Claim 2: “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.” 260

d. Claim 3: “The method of claim 1, wherein the identifier site is sequenced with the indexing site.” 261

e. Claim 5: “The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.” 263

f. Claim 6: “The method of claim 5, wherein the nucleic acid fragments from each sample have the same indexing site, and each sample has a unique indexing site.” 264

g. Claim 7: “The method of claim 6, wherein the sequencing reads are separated based on the indexing site.” 265

h. Claim 8: "The method of claim 7, wherein the separation of sequencing reads is performed prior to step d)."....266

i. Claim 9: "The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments."....267

j. Claim 10: "The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."....268

k. Claim 11: "The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length."....268

l. Claim 14: "The method of claim 1, wherein the identifier site is 8 nucleotides in length."....270

m. Claim 15: "The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site."....273

n. Claim 16: "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site."....273

o. Claim 17: "The method of claim 6, wherein the method further comprises separating fragments or sequencing reads based on indexing site."....274

3. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The '357 Patent....275

a. Bielas Anticipates Claim 1....276

1) "A method for detecting duplicate sequencing reads, the method comprising:"....276

2) "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”.....277

3) “amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences” 279

4) “and identifying sequence reads with identical identifier and target sequences as duplicates.” 280

b. To The Extent Bielas Does Not Anticipate Claim 1, Claim 1 Would Have Been Obvious Based On Bielas. 281

c. Claim 2: “The method of claim 1, further comprising removing a duplicate read from the sequence reads.” 282

d. Claim 3: “The method of claim 1, wherein the identifier site is sequenced with the target sequence.” 282

e. Claim 4: “The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.” 284

f. Claim 6: “The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.” 286

g. Claim 7: “The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site.” 287

h. Claim 9: “The method of claim 1, wherein each unique identifier is 8 nucleotides in length.” 288

i. Claim 10: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic acid fragment." 289

4. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The '241 Patent. 292

a. Bielas Anticipates Claim 1. 292

1) "A method for detecting a duplicate sequence read from a population of sample sequencing reads" 293

2) "obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor" 293

3) "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment" 295

4) "sequencing the amplicons to generate sequence reads that include identifier and target sequences; and" 296

5) "identifying sequence reads with identical identifier and target sequences as duplicates." 297

b. Claim 2: "The method of claim 1, further comprising removing a duplicate read from the sequence of reads." 298

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence." 299

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence." 299

e. Claim 5: "The method of claim 1, wherein the amplicons include amplified fragments of cDNA."....300

f. Claim 6: "The method of claim 1, wherein the amplicons include amplified fragments of genomic DNA."....301

g. Claim 7: "The method of claim 1, wherein each adaptor includes a primer binding site."....301

h. Claim 8: "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site."....302

i. Claim 9: "The method of claim 1, wherein each unique identifier is at least 6 nucleotides in length."....302

j. Claim 10: "The method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction."....303

k. Claim 11: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment."....303

l. Claim 12: "The method of claim 1, further comprising determining the genomic start position for each sequencing read."....304

m. Claim 13: "The method of claim 1, wherein each adaptor further comprises an indexing site."....305

n. Claim 14: "The method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length."....306

o. Claim 15: "The method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and

performing an amplification reaction to obtain the amplicons.” 306

p. Claim 16: “The method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index.” 307

VII. PRIOR ART CONSIDERED DURING PROSECUTION.....309

VIII. CONCLUSION.....310

1. My name is Peter Sims. I have been retained by Petitioner Integrated DNA Technologies, Inc. ("IDT") as an expert witness in this *inter partes review* ("IPR").

2. I have been asked to consider whether certain claims of patents assigned to patent owner Tecan Genomics, Inc. ("Tecan") would have been novel over the prior art or obvious to the person of ordinary skill in the art. Specifically, this declaration sets forth my opinions regarding the following patents and claims:

- U.S. Patent No. 9,546,399: Claims 1–3, 5–11, and 14–17.
- U.S. Patent No. 11,098,357: Claims 1–4, 6–7, and 9–10.
- U.S. Patent No. 11,725,241: Claims 1–16.

3. My opinions in this declaration are based on my knowledge and experience, and the materials I have considered, as discussed in this declaration.

I. BACKGROUND AND QUALIFICATIONS

4. I am a tenured Associate Professor in the Department of Systems Biology and in the Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians & Surgeons at Columbia University. I also currently serve as the Faculty Director of the Columbia Single Cell Analysis Core in the Department of Systems Biology, Vagelos College of Physicians & Surgeons at Columbia University, as the Scientific Director of the Sulzberger Columbia Genome Center, Vagelos College of Physicians & Surgeons at Columbia University, as Co-Leader

of the Precision Oncology and Systems Biology Program in the Herbert Irving Comprehensive Cancer Center, Vagelos College of Physicians & Surgeons at Columbia University, and as a Member of the Planning Committee for the Chan Zuckerberg Biohub New York. I have 20 years of experience in chemistry and systems biology.

5. I received my Ph.D. in Chemistry at Harvard University in 2009. I then pursued postdoctoral studies at Harvard University from 2009 until 2011. During my training at Harvard, I developed next-generation sequencing technology, including novel sequencing chemistry, instrumentation, and library construction methodology. The latter included methods for incorporating unique molecular identifier (UMI) barcodes into sequencing libraries for mitigating artifacts that occur during PCR amplification and sequencing.

6. I began as an Assistant Professor in the Department of Systems Biology and in the Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians & Surgeons at Columbia University in 2012. I served as an Assistant Professor until 2020, when I was promoted to Associate Professor. Since 2015, I have co-taught a class on deep sequencing covering the history and current approaches in the field of next generation sequencing. The course covers topics such as the technical details underlying next-generation sequencing instrumentation, library construction methodology for various sequencing assays (including the use

of unique molecular identifiers), and computational methods for analyzing sequencing data.

7. The focus of my work has varied over time, but has always revolved around systems biology and genomic/nucleic acid sequencing. My work has advanced the understanding of oncologic malignancies, neurological disorders, and immunology through the development and use of experimental and computational methods. These methods have been used for single-cell, cell type-specific, and whole-genome analysis of samples from patients affected by various oncologic malignancies and neurological disorders. Much of my work involves the development and application of methods for sequencing RNA and DNA from individual cells. Because individual cells contain very little RNA and DNA, methods for nucleic acid amplification, like PCR, are critically important. However, PCR can result in artifacts that reduce the accuracy of nucleic acid quantification and mutational analysis. To mitigate these artifacts, I have frequently developed and employed library construction and computational methods for single cell analysis that incorporate UMIs.

8. I have published over 95 articles in peer-reviewed journals, of which a significant majority employ systems biology and sequencing approaches to elucidate the complexities of cancer, neurologic disorders, and immunology. Several of those publications address next generation sequencing methods or cover experiments

where next generation sequencing methods were used. Such publications include:

- **Sims, P.A.**, Greenleaf, W.J., Duan, H., Xie, X.S. Fluorogenic DNA Sequencing in PDMS Microreactors. Nature Methods, 8, 575–580 (2011).

- Shiroguchi, K., Jia, T.Z., **Sims, P.A.**, Xie X.S. Digital RNA Sequencing Minimizes Sequence-Dependent Bias and Amplification Noise with Optimized Single-Molecule Barcodes. *Proc. Natl. Acad. Sci. USA*, 109, 1347–1352 (2012).

- Men, Y., Fu, Y., Chen, Z., **Sims, P.A.**, Greenleaf, W.J., Huang, Y. Digital Polymerase Chain Reaction in an Array of Femtoliter Polydimethylsiloxane Microreactors. *Analytical Chemistry*, 84, 4262–4266 (2012).

- Gonzalez, C., Sims, J.S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P., **Sims. P.A.** Ribosome Profiling Reveals a Cell-Type-Specific Translational Landscape in Brain Tumors. *J. Neuroscience*, 34, 10924–10936 (2014).

- Gill, B.J., Pisapia, D.J., Malone, H.R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J.S., Banu, M., Dovas, A., Teich, A.F., Sheth, S., McKhann, G.M., Sisti, M.B., Bruce, J.N., **Sims, P.A.**, Canoll, P. MRI-Localized Biopsies Reveal Subtype-Specific Differences in Molecular and Cellular Composition at the Margins of Glioblastoma. *Proc. Natl. Acad. Sci. USA*, 111, 12550–12555 (2014).

- Bose, S., Wan, Z., Carr, A., Rizvi, A.H., Vieira, G., Pe'er, D., **Sims, P.A.** Scalable Microfluidics for Single-Cell RNA Printing and Sequencing. *Genome Biology*, 16, 120 (2015).

- Hornstein, N., Torres, D., Das Sharma, S., Tang, G., Canoll, P., **Sims, P.A.** Ligation-free ribosome profiling of cell type-specific translation in the brain, *Genome Biology*, 17, 149 (2016).

- Yuan, J. and **Sims, P.A.** An automated microwell platform for large-scale single cell RNA-Seq, *Scientific Reports*, 6, 33883, (2016).

- Bush., E.C., Ray, F., Alvarez, M.J., Realubit, R., Li, H., Karan, C., Califano, A., **Sims, P.A.** PLATE-Seq for Genome-Wide Regulatory Network Analysis of High-Throughput Screens. *Nature Communications*, 8, 105 (2017).

- Yuan, J., Sheng, J., **Sims, P.A.** SCOPE-seq: A Scalable Technology for Linking Live Cell Imaging and Single-Cell RNA-seq. *Genome Biology*, 19, 227 (2018).

- Levitin, H.M., Yuan, J., Cheng, Y.L., Ruiz, F.J.R., Bush, E.C., Bruce, J.N., Canoll, P., Iavarone, A., Lasorella, A., Blei, D.M., **Sims, P.A.** *De novo* Gene Signature Identification from Single-Cell RNA-seq with Hierarchical Poisson Factorization. *Molecular Systems Biology*, 15, e8557 (2019).

- Szabo, P.A., Levitin, H.M., Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., Farber, D.L., **Sims, P.A.** Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, 10, 4706 (2019).

- Liu, Z., Yuan, J., Lasorella, A., Iavarone, A., Bruce, J.N., Canoll, P., **Sims, P.A.** Integrating single-cell RNA-seq and imaging with SCOPE-seq2. *Scientific Reports*, 19482 (2020).

- Zhao, W., Dovas, A., Spinazzi, E.F., Levitin, H.M., Banu, M.A., Upadhyayula, P., Sudhakar, T., Marie, T., Otten, M.L., Sisti, M.B., Bruce, J.N., Canoll, P., **Sims, P.A.** Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. *Genome Medicine*, 13, 82 (2021).

- Metz, J.B., Hornstein, N.J., Das Sharma, S., Worley, J., Gonzalez, C., **Sims, P.A.** High-throughput translational profiling with riboPLATE-seq. *Scientific Reports*, 12, 5718 (2022).

- Cheng, Y.L., Banu, M.A., Zhao, W., Rosenfeld, S.S., Canoll, P., **Sims, P.A.** Multiplexed single-cell lineage tracing of mitotic kinesin inhibitor resistance in glioblastoma. *Cell Reports*, 43, 114139 (2024).

9. I am also a named inventor on fourteen patents that have resulted from

my research. The majority of the patents on which I am a named inventor relate to next generation sequencing technologies and methods like those implicated in this declaration, including the use of unique molecular tags used to identify copies of original molecules from a sample after sequencing. Such patents include:

- Xie X.S., Shiroguchi K., **Sims P.A.**, Jia T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 9,260,753 (2016).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 10,287,630 (2019).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 10,584,382 (2020).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,035,001 (2021).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,078,533 (2021).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,286,523 (2022).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,352,669 (2022).
- Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,608,527 (2023).
- Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,629,379 (2023).
- Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,834,712 (2023).
- **Sims, P.A.**, Bose, S., Yuan, J. RNA Printing and Sequencing Devices, Methods, and Systems. U.S. Patent 11,788,120 (2023).

- Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,866,781 (2024).

10. In addition to my research and teaching, I have also served as a reviewer in numerous scientific journals, including Cell Genomics, Cell Systems, Genome Biology, Genome Medicine, Genome Research, Molecular Systems Biology, and Nature Genetics, among others.

11. I have served as a member of the editorial board for Genome Medicine since 2022 and as a Guest Editor for Current Opinion in Systems Biology, Genetics and Epigenomics in 2018.

12. Throughout my career, I have received numerous awards for my work, including the Research Initiatives for Science and Engineering (RISE) Award from Columbia University in 2021 and the Emerging Leader Award from The Mark Foundation for Cancer Research in 2018.

13. I have given over 60 invited presentations on my research, both in the U.S. and internationally, including at leading research institutes including Cold Spring Harbor Laboratory, Yale University School of Medicine, Johns Hopkins University, Columbia University, and the National Cancer Institute, among others.

14. My experience and qualifications are further set forth in my curriculum vitae, attached hereto as Appendix A.

15. I am being compensated at the rate of $450 per hour, plus expenses.

My compensation is in no way dependent on the outcome of this proceeding.

16. I have testified at trial or deposition in one prior matter, *Scale Biosciences, Inc. v. Parse Biosciences, Inc.*, No. 1:22-CV-01597 (D. Del.).

II. MATERIALS CONSIDERED

17. In formulating the opinions in this declaration, I have relied upon my decades of experience, training, and knowledge regarding bioinformatics, genomics, and nucleic acid sequencing. In addition, I considered the documents cited in this declaration, which I understand have been submitted as exhibits in conjunction with this IPR.

18. For reference, I have provided below a table of the references that I have cited herein, as well as their respective exhibit numbers:

Exhibit	Short Name	Document
Ex. 1003	’399 patent	U.S. Patent No. 9,546,399
Ex. 1004	’357 patent	U.S. Patent No. 11,098,357
Ex. 1005	’241 patent	U.S. Patent No. 11,725,241
Ex. 1007	Siebert	Siebert et al., *An improved PCR method for walking in uncloned genomic DNA*, NUCLEIC ACIDS RES. 23(6) 1087–1088 (1995)

Exhibit	Short Name	Document
Ex. 1008	Caruccio	Caruccio, *Preparation of Next-Generation Sequencing Libraries using Nextera™ Technology: Simultaneous DNA Fragmentation and Adaptor Tagging by In Vitro Transposition*, HIGH-THROUGHPUT NEXT GENERATION SEQUENCING: METHODS AND APPLICATIONS, METHODS IN MOLECULAR BIOLOGY, Ch. 17 (2011)
Ex. 1009	Bronner	Bronner et al., *Improved Protocols for Illumina Sequencing*, CURR PROTOC HUM GENET. (2009)
Ex. 1011	Iafrate	U.S. Patent Application Publication No. 2013/0303461
Ex. 1012	Kivioja	Kivioja et al., *Counting absolute numbers of molecules using unique molecular identifiers*, NATURE METHODS (2011)
Ex. 1013	Bielas	International Publication No. WO 2013/123442 A1
Ex. 1014	Masoudi-Nejad	Masoudi-Nejad et al., *Next Generation Sequencing and Sequence Assembly Methodologies and Algorithms* (2013)
Ex. 1015	Zheng	Zheng et al., *Anchored multiplex PCR for targeted next-generation sequencing*, NATURE MEDICINE (2014)
Ex. 1018		File History for U.S. Patent No. 9,546,399
Ex. 1019		File History for U.S. Patent No. 11,098,357
Ex. 1020		File History for U.S. Patent No. 11,725,241

Exhibit	Short Name	Document
Ex. 1021	Miner	Miner et al., *Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR*, NUCLEIC ACID RESEARCH 32:17 (2004)
Ex. 1022	McCloskey	McCloskey et al., *Encoding PCR Products with Batch-stamps and Barcodes*, BIOCHEM GENET 45:761–767 (2007)
Ex. 1023	Hiatt	Hiatt et al., *Parallel, tag-directed assembly of locally derived short sequence reads*, NAT METHODS 7(2):119–122 (2010)
Ex. 1024	König	König et al., *iCLIP reveals the function hnRNP particles in splicing at individual nucleotide resolution*, NAT STRUC MOL BIOL. 17(7):909–915 (2010)
Ex. 1025	Kinde	Kinde et al., *Detection and quantification of rare mutations with massively parallel sequencing*, PNAS 108(23):9350–9535 (2011)
Ex. 1026	Casbon	Casbon et al., *A method for counting PCR template molecules with application to next-generation sequencing*, NUCLEIC ACIDS RESEARCH 39:12 (2011)
Ex. 1027	Fu	Fu et al., *Counting individual DNA molecules by the stochastic attachment of diverse labels*, PNAS 108(22):9026–9031 (2011)
Ex. 1028	Jabara	Jabara et al., *Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID*, PNAS 108(50):20166–20171 (2011)

Exhibit	Short Name	Document
Ex. 1029	Tollervey	Tollervey et al., *Characterising the RNA targets and position-dependent splicing regulation by TDP-43; implications for neurodegenerative diseases*, NAT NEUROSCI. 14(4):452–458 (2011)
Ex. 1030	Shiroguchi	Shiroguchi et al., *Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes*, PNAS 109(4):1347–1352 (2012)
Ex. 1031	Schmitt	Schmitt et al., *Detection of ultra-rare mutations by next-generation sequencing*, PNAS 109(36):14508–14513 (2012)
Ex. 1032	De Luca	De Luca and Dono, *The Opportunities and Challenges of Molecular Tagging Next-Generation Sequencing in Liquid Biopsy*, MOLECULAR DIAGNOSIS & THERAPY 25:537–547 (2021)
Ex. 1033	Samuels	U.S. Patent Application Publication No. 2012/0220494
Ex. 1050	Moorthie	Moorthie et al., *Review of massively parallel DNA sequencing technologies*, HUGO JOURNAL 5:1–12 (2011)
Ex. 1053	Syed	Syed, *Chapter 19: Second-Generation Sequencing Library Preparation:* In Vitro *Tagmentation via Transposome Insertion*, *in* TAG-BASED NEXT GENERATION SEQUENCING (Matthias Harbers and Günter Kahl, eds., 2011)

Exhibit	Short Name	Document
Ex. 1054	Loeb	Loeb et al., *Transcriptome-wide miR-155 Binding Map Reveals Widespread Noncanonical MicroRNA Targeting*, MOLECULAR CELL 760–770 (2012)
Ex. 1055	Kircher	Kircher et al., *Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform*, NUCLEIC ACIDS RESEARCH 40:1 (2012)
Ex. 1056	Illumina Letter	Illumina Customer Letter (Sept. 7, 2012)
Ex. 1057	Nextera Prep Guide	Nextera® XT DNA Sample Preparation Guide (Oct. 2012)
Ex. 1058	Troutt	Troutt et al., *Ligation-anchored PCR: A simple amplification technique with single-sided specificity*, PNAS, USA 9823–9825 (1992)
EX. 1059	Hug & Schuler	Hug and Schuler, *Measurement of the Number of Molecules of a Single mRNA Species in a Complex mRNA Preparation*, J. THEOR. BIOL. 221:615–624 (2003),
Ex. 1060	Meyer & Kircher	Meyer and Kircher, *Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing*, COLD SPRING HARBOR PROTOCOLS (2010)
Ex. 1061	Kircher Thesis	Kircher Thesis, *Understanding and improving high-throughput sequencing data production and analysis*, LEIPZIG UNIVERSITY, Germany (2011)

Exhibit	Short Name	Document
Ex. 1062	Smith Thesis	Smith Thesis, *Low-Level Variant Detection in Human Mitochondrial DNA Using The Illumina® MiSeq™ Next-Generation Sequencing (NGS) Platform*, WESTERN CAROLINA UNIVERSITY (2013)
Ex. 1063	Won	Won et al., *Detecting Somatic Genetic Alterations in Tumor Specimens by Exon Capture and Massively Parallel Sequencing*, J. VISUALIZED EXPERIMENTS (2013)
Ex. 1064	Kozich	Kozich et al., *Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform*, APPLIED AND ENVIRONMENTAL MICROBIOLOGY 79:17, 5112–5120 (2013)
Ex. 1065	Pray	Pray, *Major Molecular Events of DNA Replication*, NATURE EDUCATION (2008)
Ex. 1066	Mackenzie	Mackenzie, *DNA vs. RNA – 5 Key Differences and Comparison*, TECHNOLOGY NETWORKS (2020)
Ex. 1067	Clancy	Clancy and Brown, *Translation: DNA to mRNA to Protein*, NATURE EDUCATION (2008)
Ex. 1068	Gauthier Thesis	Gauthier Thesis, *Simulation of polymer translocation through small channels*, UNIVERSITY OF OTTAWA, Canada (2008)
Ex. 1069	Lohman	Lohman et al., *DNA Ligases*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 3.14.1-3.14.7 (2011)

Exhibit	Short Name	Document
Ex. 1070	Margulies	Margulies et al., *Genome sequencing in microfabricated high-density picolitre reactors*, NATURE 376–380 (2005)
Ex. 1071		DNA Sequencing Costs: Data, https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
Ex. 1072	Craig	Craig et al., *Identification of Genetic Variants Using Barcoded Multiplexed Sequencing*, NAT. METHODS 5(10) 887–893 (2008)
Ex. 1073	Parameswaran	Parameswaran et al., *A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing*, NUCLEIC ACIDS RESEARCH 35:19 (2007)
Ex. 1074	Lonigro	Lonigro et al., *Detection of Somatic Copy Number Alterations in Cancer Using Targeted Exome Capture Sequencing*, NEOPLASIA 13(11) 1019–1025 (2011)
Ex. 1075	Bielas Annotated Figure	Annotated Figure
Ex. 1076	Bielas Illumina P5	Annotated Figure
Ex. 1077	Bielas Read 1 Primer	Annotated Figure
Ex. 1078	Bielas i5 Index	Annotated Figure

III. LEGAL STANDARDS

A. Burden of Proof

19. I understand that in an *inter partes* review, the Petitioner, IDT, has the burden of proving invalidity by a "preponderance of the evidence," which I understand means that IDT must show that it is more likely than not that the patent is invalid.

B. Anticipation And Obviousness Are Evaluated From The Perspective Of The POSA.

20. I have been instructed by counsel that the question of whether a claim is anticipated or obvious based on the prior art is evaluated from the perspective of a hypothetical person of ordinary skill in the art (which I refer to in this declaration as "the POSA"). I understand that the POSA is a hypothetical individual with ordinary skill in the relevant field who is deemed to have knowledge of the relevant prior art at the priority date of the patent. I have provided my opinions regarding the background and qualifications of the POSA in Section V of this declaration below.

C. Anticipation

21. I have been instructed by counsel that a claim is invalid as anticipated if a single prior art reference discloses every limitation of the claim as arranged in the claim. I have been informed by counsel that, for a reference to anticipate the claim, the reference must provide disclosure sufficient for the POSA to practice the relevant disclosure without undue experimentation, but that the reference need not

include details that are already well-known to the POSA.

22. I understand from counsel that the limitations can be disclosed either expressly or inherently or implicitly in a reference. In this regard, I understand that the question for anticipation is what the POSA would understand based on the teachings of the prior art reference, and not what the authors or inventors of the prior art actually did in the laboratory. A limitation is inherently or implicitly disclosed if it is a natural result flowing from the explicit disclosure of the prior art. I further understand that a particular feature can be inherently anticipated even if it was not recognized by those of skill in the art at the time.

23. I understand that a prior art species falling within a claimed genus will anticipate the genus. Likewise, I understand that a specific value disclosed in the prior art will anticipate a later claimed range of values encompassing the specific disclosed value.

24. I further understand from counsel that a prior art reference can anticipate if it discloses a sufficiently small group of combinations such that the POSA could at once envisage the claimed combination from within that group.

D. Obviousness

25. I have been informed by counsel that a patent claim is invalid if the claimed invention would have been obvious to the POSA at the time the alleged invention was made. Whether a claimed invention is obvious depends on several

factors, including (1) the scope and content of the prior art, (2) the differences, if any, between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) objective indicia of nonobviousness, if any. I understand from counsel that for an invention to be obvious, the POSA must have had a reason or been motivated to modify the prior art, or to combine one or more prior art references, in a manner that would yield the claimed invention.

26. I have been informed by counsel that where the prior art discloses a range of values for a particular variable, and the determination of that particular variable is a matter of routine optimization for the POSA, there is a presumption that it would have been obvious to arrive at a particular value within the range. I understand that this presumption can be overcome by a showing that the prior art taught away from a particular value or portion of the prior art range, or that the claimed value or portion of the prior art range displays unexpected properties.

27. I have further been instructed by counsel that where the general conditions of a claim are disclosed in the prior art, it is not inventive to discover the optimum or workable ranges by routine experimentation. I further understand that a variable is "result effective" if there was a recognition in the prior art that a property is affected by a particular claimed variable. When a variable is result effective, there is a presumption that the routine optimization of that variable for that property is obvious.

28. I have also been informed by counsel that for a claim to be obvious the POSA must have had a reasonable expectation of success in making the relevant modifications or combining the prior art. I understand that a reasonable expectation of success does not require absolute certainty of success.

29. I have also been instructed by counsel that objective indicia (also called secondary considerations) may support nonobviousness. I understand that these objective indicia may include copying, long felt but unmet need, failure of others, commercial success, unexpected results created by the claimed invention, unexpected properties of the claimed invention, and skepticism of skilled artisans before the invention. However, I further understand that these objective indicia can only support nonobviousness where the objective indicia have a nexus to the claim, which requires a showing that the objective indicia result from what is claimed.

IV. TECHNICAL BACKGROUND

30. The ’399, ’357, and ’241 patents (the “identifier” patents) generally relate to a method of tagging DNA molecules with an “identifier” sequence prior to NGS sequencing in order to determine which “duplicate” sequencing reads are true duplicates from the original sample, and which are copies of a single original molecule from the original sample that were generated during library preparation.

A. DNA

31. DNA (deoxyribonucleic acid) is one of several nucleic acids, which are

biomolecules comprised of a chain of individual units of nucleotides that I discuss further below. Relatively short strands of nucleic acids are sometimes referred to as "oligonucleotides."

32. DNA encodes genetic information for the development and function of almost all living organisms, including humans. DNA is, in most cases, a double-helix-shaped polymer composed of two "polynucleotide strands" coiled together. However, DNA can also sometimes exist as a single-stranded molecule. The "polynucleotide" strands are so-named because each is made up of many "nucleotides" held together by chemical bonds. Each nucleotide is composed of (1) one of four "base" groups, (2) a common sugar (2-deoxyribose) group, and (3) a phosphate group. The four bases found in DNA are adenine ("A"), cytosine ("C"), guanine ("G"), and thymine ("T"). Each base group has a unique structure, as shown in the figure below, which I have reproduced from EX1014 (Masoudi-Nejad) at 0011.

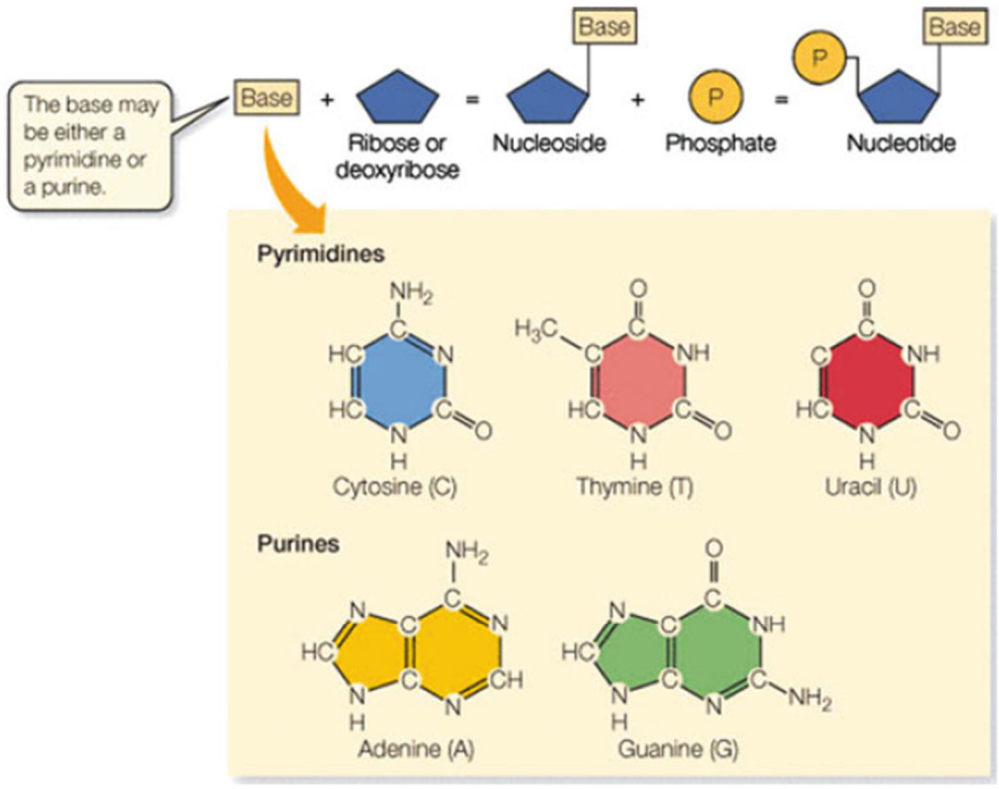


Fig. 1.1 Three components of each nucleotide: the nitrogenous base that can basically belong to two categories (*single ring*: pyrimidines, or *two-linked rings*: purines), a pentose sugar (ribose in RNA and deoxyribose in DNA), and a phosphate group [2]

33. Sequences of adenine, cytosine, guanine, and thymine are repeated throughout DNA in both "coding" and "noncoding" regions. Coding regions contain instructions for building cellular proteins; noncoding regions do not contain such instructions but may serve other functions, such as regulating gene expression. DNA sequences are specific and control biological functions and the expression of an organism's phenotype.

34. A DNA strand is generated from the incorporation of individual nucleotide units called deoxyribonucleotide triphosphates or "dNTPs." Each dNTP

is composed of the base, sugar, and a substituted phosphate group that is bonded to two additional phosphate groups. When a new dNTP is added to a DNA strand, two phosphate groups from the dNTP are lost as part of the reaction. Each sugar ring in DNA is substituted at the same positions with each group. Specifically, the base is always found at the 1′ position on the ring, the phosphate group is always found at the 5′ position, and a hydroxyl group is always found at the 3′ position.

35. Each strand of a DNA molecule is joined together from individual nucleotides in a specific arrangement. The phosphate and sugar groups of each nucleotide are covalently bonded together to form each strand's structural "backbone." The 5′ phosphate group of one nucleotide is always bound to the 3′ hydroxyl group of the preceding nucleotide in the chain. Thus, because of the particular chemical attachment and the asymmetrical nature of each nucleotide, each strand of a DNA molecule has a specific directionality from the 5′ phosphate of the first nucleotide through to the 3′ hydroxyl of the last nucleotide in the chain. EX1014 (Masoudi-Nejad) at 0011-12.

36. In a double stranded DNA molecule, the two strands of DNA are held together (or "hybridized" or "annealed" to each other) (1) via noncovalent interactions between the bases on opposite strands, specifically hydrogen bonds, and (2) due to favorable stabilization interactions within each strand, specifically base-stacking interactions between the bases on the strand. Because of the specific

structures of the individual base groups found in each nucleotide, each base can only form sufficiently favorable interactions to hybridize to one other base. Specifically, "A" pairs with "T" and "G" pairs with "C" to form A-T and G-C "base pairs." Accordingly, each base in a base pair, and each strand in a double stranded molecule of DNA, is said to be "complementary" to the other. The size of the bases (fused bicyclic rings for the purine bases adenine and guanine vs. single rings for the pyrimidine bases cytosine and thymine) and the particular position of hydrogen bond donors and acceptors within each base ensure the specific complementarity of each base pair and inhibit nonspecific (mismatched) hybridization. The figure below, which I have reproduced from EX1014 (Masoudi-Nejad) at 0012, depicts the complementary relationship between bases and DNA's double-stranded, helical structure.

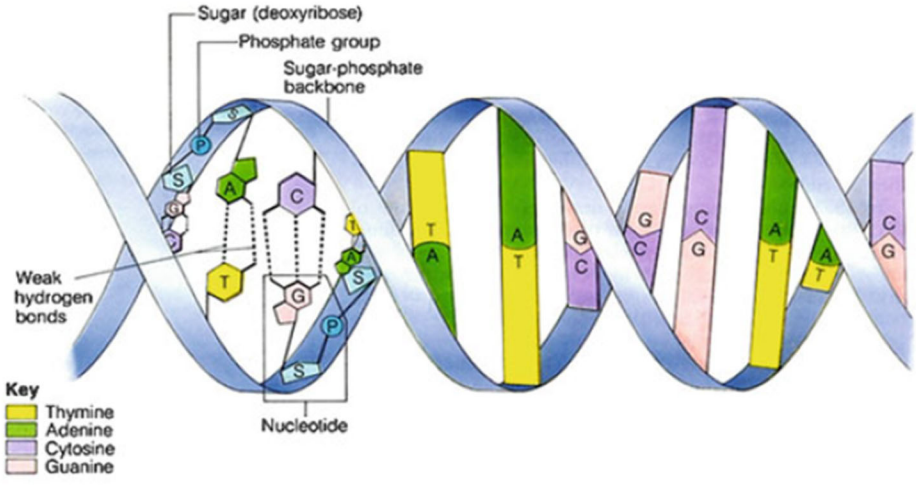


Fig. 1.3 Double-helical structure of DNA. The chains of sugar-phosphate groups are linked together by complementary bases [2]

37. The order in which the base-pairs appear in the DNA molecule is referred to as the DNA "sequence," with each individual strand having a sequence that is "complementary" to the sequence of the other strand, due to the A-T and G-C base pairing. As I discuss further below, genetic information that controls the physical expression of traits in an organism is encoded in the specific "sequence" of the nucleotides in a strand of DNA. The complementarity of the DNA strands allows for the replication of that genetic information from one strand of DNA to another strand, thus allowing for the propagation and replication of DNA during cell division.

38. DNA strands and their sequences are conventionally shown to run from 5′ to 3′ when reading left to right, and with the strand running 5′ to 3′ on "top" in a double-stranded depiction. In this declaration, for simplicity and clarity, I sometimes refer to the strand that runs 5′ to 3′ is referred to as the "sense" strand, and the strand that runs 3′ to 5′ as the "antisense" strand. Consistent with this convention, throughout this declaration, simplified diagrams are used to depict DNA strands as: (1) two lines, as shown by "(a)" in the exemplary figure below, or, (2) where information about the strand's "sequence" is relevant, as two strings of text, with each base represented by its corresponding letter, "(b)" in the exemplary figure below, or (3) some combination of (1) and (2). Where applicable, the diagrams included in this declaration also denote the 5′ and/or 3′ ends of the individual strands,

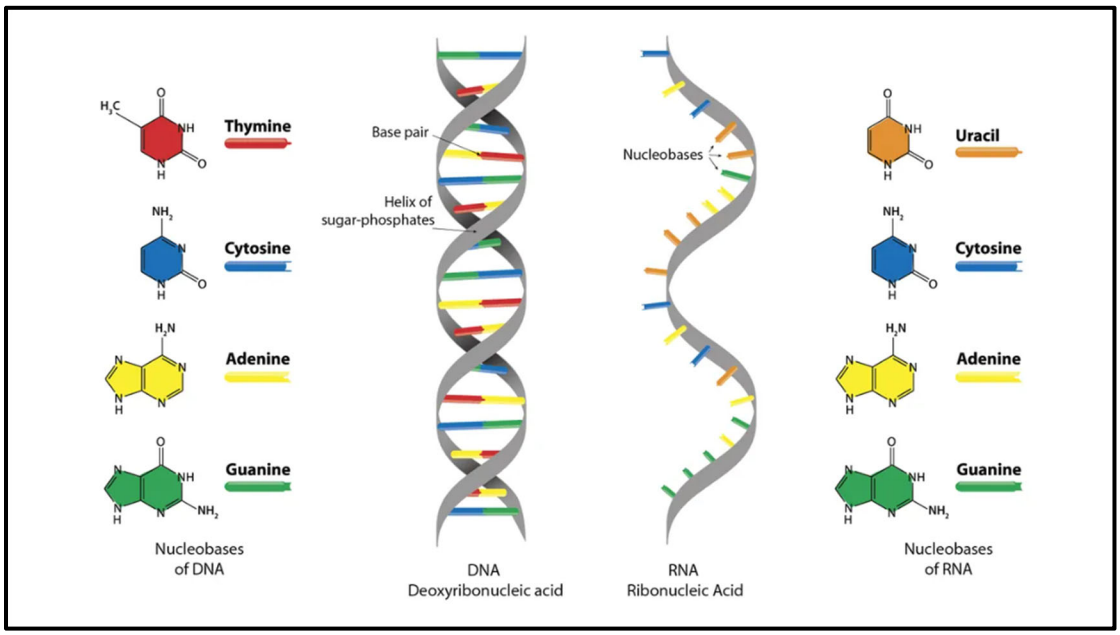
as shown below.



B. RNA

39. Another type of nucleic acid that, like DNA, contains genetic information, is ribonucleic acid ("RNA"). RNA is structurally very similar to DNA, in that it is made up of nucleotides, but with two notable exceptions: *First*, the sugar-phosphate "backbone" of RNA contains a ribose sugar in place of the deoxyribose sugar in DNA. *Second*, RNA uses uracil ("U"), an unmethylated form of thymine, as the base complementary to "A," in place of thymine (T) in DNA.

40. Unlike DNA, which is generally found in a double-stranded form, most RNA molecules in mammalian cells exist in single-stranded form. There are several types of RNA that are most commonly found in cells, including ribosomal RNA ("rRNA"), transfer RNA ("tRNA"), and messenger RNA ("mRNA"). The figure below, which I have reproduced from EX1066 (Mackenzie) at 0001, depicts the differences in bases and structure between RNA and DNA.



41. Like DNA, RNA serves a critical biological function. Specifically, RNA helps translate genetic information into proteins. mRNA is transcribed from DNA and thus carries the genetic information necessary for synthesizing proteins. rRNA plays crucial roles in both the structure and catalytic function of ribosomes, which are cellular organelles necessary for protein synthesis. The ribosome translates mRNA into proteins by incorporating amino acids carried by tRNA into a growing polypeptide chain. The polypeptide chain is folded into the functional protein.

C. DNA Synthesis

42. Although DNA is usually double-stranded, it is synthesized one strand at a time, using an already-formed single-stranded molecule as a "template." The generation of a new strand with a sequence complementary to the template strand's

sequence begins by annealing a "primer"—that is, a shorter single-stranded molecule with a sequence complementary to a sequence on the template strand—to the template strand. In nature, primers are frequently composed of short stands of RNA, but in a test tube, scientists often use short single strands of DNA as primers as discussed further below.

43. An enzyme known as a "polymerase" is used to extend the primer (always in the 5′ to 3′ direction) so as to generate a second strand with a sequence complementary to the template strand. The polymerase catalyzes a reaction in which the free 3′ hydroxyl (OH) group from the growing strand of DNA attacks the primary 5′ phosphate on the next dNTP to be added to the new strand. EX1014 (Masoudi-Nejad) at 0011-12; EX1065 (Pray) at 0001-02. The directionality of DNA synthesis is shown in the figure below, which I have reproduced from EX1014 (Masoudi-Nejad) at 0012.

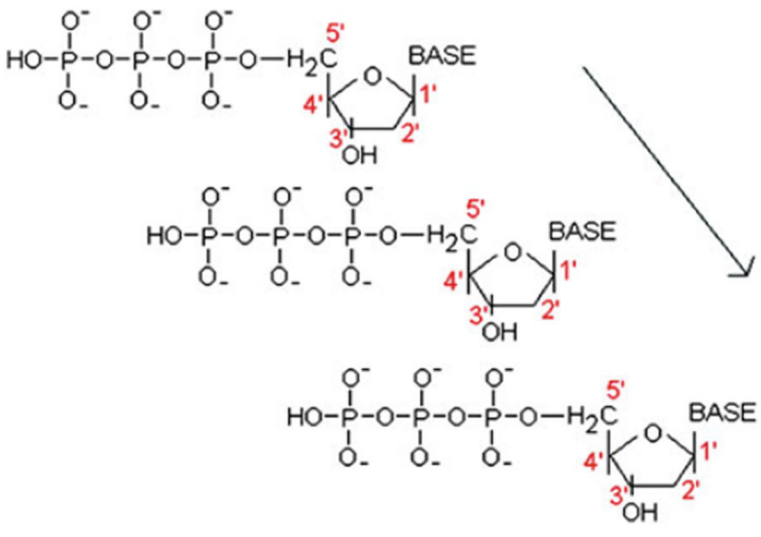


Fig. 1.4 DNA synthesis direction. The 5′ end of the new nucleotide is linked to the 3′-OH of the last nucleotide of the growing chain by DNA polymerase action. During this reaction, a pyrophosphate group is released [http://www.prism.gatech.edu/~gh19/b1510/dnarep.htm]

44. This interaction releases a "pyrophosphate" (a molecule composed of two covalently-bound phosphate groups) from the incoming dNTP, thereby allowing the "new" dNTP to form a phosphodiester bond with the growing strand of DNA. The newly added nucleotide has a free 3′-hydroxyl group that is available to repeat the reaction with the next dNTP to be added. EX1014 (Masoudi-Nejad) at 0011-12. This synthesis process is sometimes referred to as "elongation" or "extension" of a polynucleotide chain. DNA synthesis is depicted in the figure below, which I have reproduced from EX1065 (Pray) at 0002.

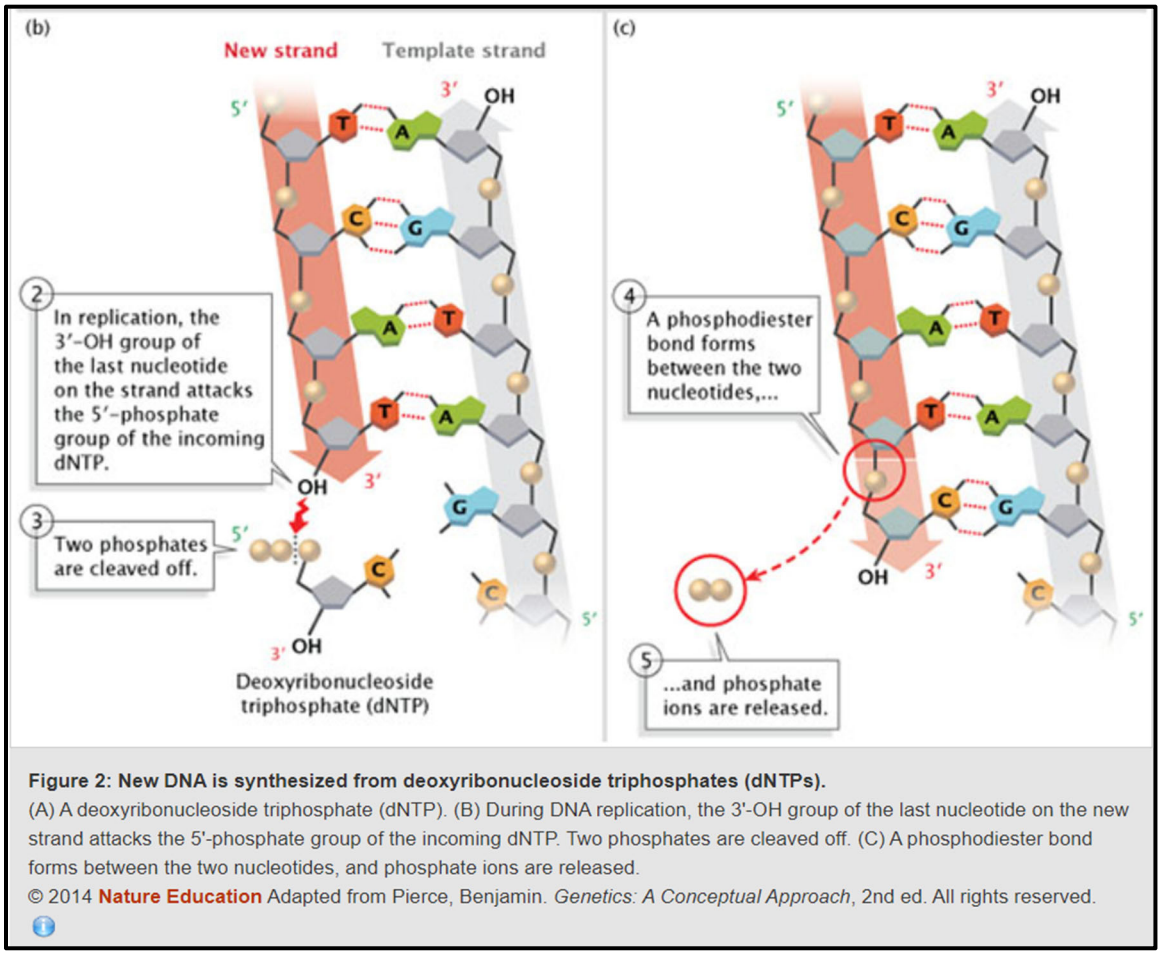


Figure 2: New DNA is synthesized from deoxyribonucleoside triphosphates (dNTPs).
(A) A deoxyribonucleoside triphosphate (dNTP). (B) During DNA replication, the 3'-OH group of the last nucleotide on the new strand attacks the 5'-phosphate group of the incoming dNTP. Two phosphates are cleaved off. (C) A phosphodiester bond forms between the two nucleotides, and phosphate ions are released.
© 2014 Nature Education Adapted from Pierce, Benjamin. *Genetics: A Conceptual Approach*, 2nd ed. All rights reserved.

45. DNA synthesis is a fundamental biological process that takes place within living organisms ("*in vivo*"), but can also be done artificially in an "*in vitro*" (lab-based) setting via a synthesis reaction containing templates, primers, dNTPs, and polymerases. In the briefest summary, scientists use a purified DNA polymerase, which is then added to denatured single stranded template DNA along with specific or random primers and dNTPs. Under appropriate conditions, the primers will anneal to the template strands and the polymerase will then bind and extend the primer strands using the free dNTPs. Thus, scientists are able to use the natural function of DNA polymerase to regenerate the complementary strand of DNA in a test tube outside of an organism.

D. Genomic DNA

46. In an organism, the genetic information that determines the physical traits of an organism is encoded in DNA. DNA performs this function by providing instructions for the creation of proteins, which are the building blocks of cells and responsible for basic biological function. EX1067 (Clancy) at 0001. A "gene" is a unit of DNA that contains information for making, or "expressing," a particular protein. EX1067 (Clancy) at 0001. At a high level, a protein is complex macromolecule composed of a "polypeptide chain" of amino acids that is folded into a complex three-dimensional structure. Protein function is determined both by the amino acid sequence within the protein's polypeptide chain and the physical

structure of the polypeptide chain after it is folded into a complex three-dimensional structure.

47. At a high level, in cells, proteins are created from DNA in a two-step process of "transcription" (conversion of DNA into RNA) followed by "translation" (creation of a protein from the RNA). A specific type of RNA is involved in this process called "messenger RNA" or mRNA. During transcription, a double-stranded DNA molecule is partially denatured to single stranded form and an enzyme known as an "RNA polymerase" facilitates the generation of a single-stranded mRNA molecule with a sequence complementary to the template DNA molecule. EX1067 (Clancy) at 0001. The resulting mRNA molecule contains a single-stranded copy of the gene containing "coding" genetic material for the organism.

48. In eukaryotes, most mRNA molecules have a unique structure with a "polyA tail" at their 3′ end, comprising a chain of repetitive "A" nucleotides in a row. The polyA tail, which can be hundreds of nucleotides in length, serves to stabilize and prevent degradation of the mRNA molecule.

49. In cells, the mRNA is then "translated" via a large macromolecular complex called a ribosome. A ribosome is a large unit comprised of RNAs and proteins that catalyzes the formation of a polypeptide chain based on the sequence of an mRNA. The ribosome functions by forming a polypeptide protein molecule based on translating "codons" (*i.e.*, three-base segments of mRNA) into

corresponding amino acids. The ribosome then adds the appropriate amino acid based on each codon to the growing polypeptide chain. The polypeptide sequences form the primary structure of a protein molecule. A simplified depiction of mRNA and polypeptide synthesis is shown in the figure below, which I have reproduced from EX1067 (Clancy) at 0001.

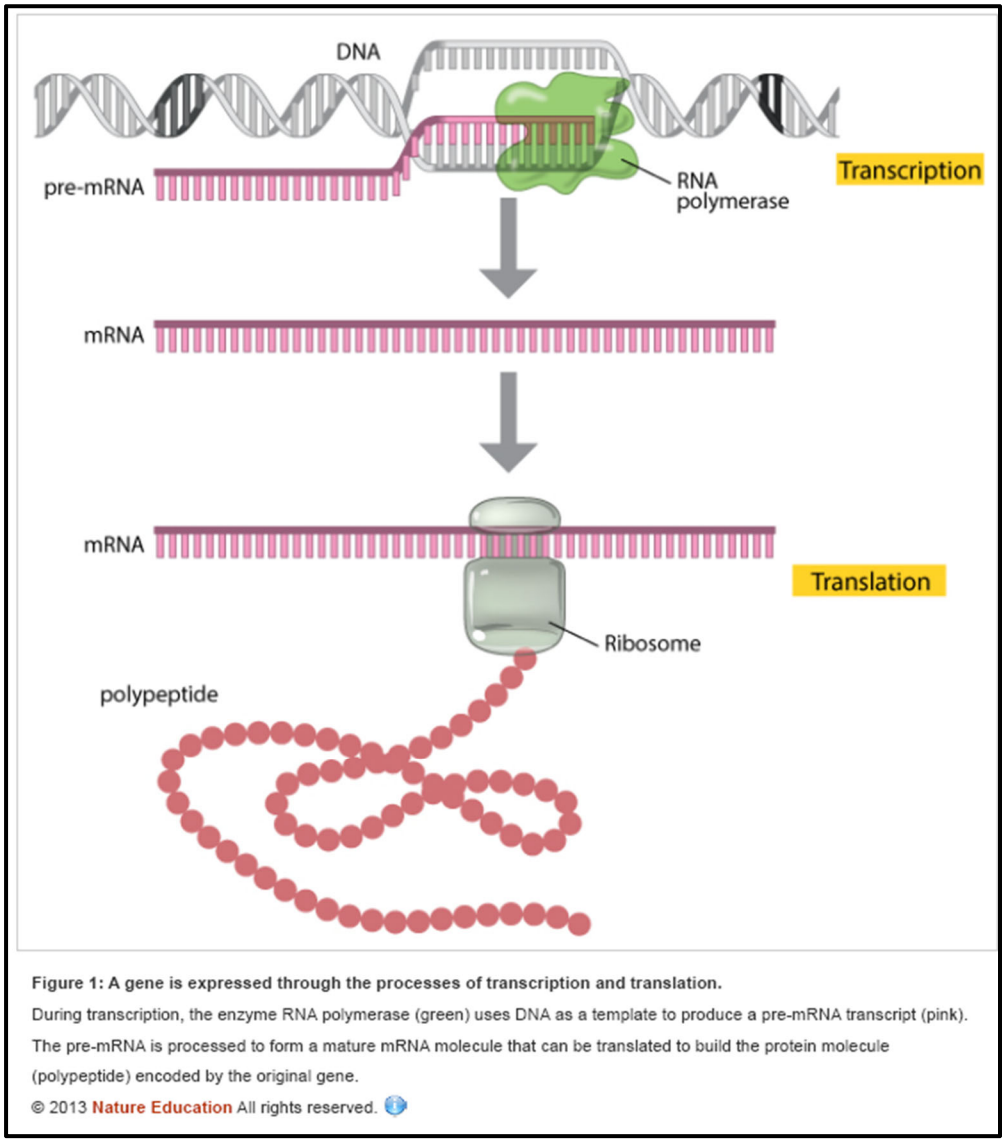


Figure 1: A gene is expressed through the processes of transcription and translation.
During transcription, the enzyme RNA polymerase (green) uses DNA as a template to produce a pre-mRNA transcript (pink). The pre-mRNA is processed to form a mature mRNA molecule that can be translated to build the protein molecule (polypeptide) encoded by the original gene.
© 2013 Nature Education All rights reserved.

50. The complete set of all of the DNA within an organism is referred to as

that organism's "genome." "Genomic DNA" refers to an organism's genetic information, including both "coding" regions that are expressed as proteins, and non-coding regions that are not expressed as proteins, but may serve regulatory or other functions.

51. The complete set of an organism's coding genetic information is reflected in the complete set of mRNA in an organism, which is known as an organism's "transcriptome."

52. RNA can be converted back into DNA using an enzyme known as "reverse transcriptase." Scientists have isolated reverse transcriptase enzymes as research tools. Scientists can thus combine an isolated reverse transcriptase with RNA and dNTPs to generate a single stranded DNA molecule complementary to a particular RNA. Such a synthetic DNA molecule is referred to as complementary DNA or cDNA. DNA polymerase can then be used to regenerate the second complementary strand of cDNA to make a double stranded cDNA molecule. Scientists use the term cDNA to generally refer to synthetic DNA generated from a particular RNA molecule.

E. Ligation and Adaptors

53. Individual DNA (or RNA) strands can also be linked together "end to end" in a process called "ligation." Linking two DNA strands together requires using an enzyme called a ligase, which was first discovered in the 1960s and used

for research purposes starting in the 1970s and 1980s. EX1069 (Lohman) at 0001. The ligase catalyzes the formation of a phosphodiester bonds between the 5′ terminal phosphate group on one strand of DNA and the 3′ terminal hydroxyl group on another strand of DNA. EX1069 (Lohman) at 0001.

54. There are different types of ligases—some catalyze the formation of a bond between two single stranded pieces of DNA, but others catalyze the joining of two double stranded pieces of DNA, thus joining the 5′ and 3′ ends of a double-stranded fragment to the corresponding 3′ and 5′ ends of another double-stranded fragment. EX1069 (Lohman) at 0001-6.

55. In order for the ligation reaction to proceed, each of the DNA fragments must have a phosphate on the 5′ terminus of the end that is to be ligated. 5′ phosphates can be added to unphosphorylated DNA using an enzyme known as a "kinase" through a process called "phosphorylation." EX1069 (Lohman) at 0001. When DNA fragments are fragmented via shearing or tearing, they often undergo an "end repair" or "end polishing" step to ensure the ends are "blunted," that is, do not contain overhanging, single-stranded segments on either end, to allow for ligation.

56. One common application of ligation is to link a naturally occurring DNA fragment to a synthetic piece of DNA. The synthetic DNA is typically a short strand of DNA called an "adaptor" that contains known sequences, some of which I discuss below in Section IV.K.2 of my declaration. Adaptors can be used in research

processes, such as DNA sequencing and anchored multiplex PCR, both of which are discussed further below.

F. DNA sequencing

57. "DNA sequencing," that is, the process of determining the sequence of a DNA strand, is a powerful technique used to obtain genetic information about an organism. DNA sequencing has long been an important tool for researchers across many disciplines because understanding genetic sequences allows insight into biological processes. Indeed, "[k]nowledge of the DNA sequence of an organism is the key to understanding how that organism functions." EX1009 (Bronner) at 0001. So beginning in the second half of the 20th century, scientists began using naturally occurring proteins, such a DNA polymerases, RNA transcriptases, and ligases, as research tools. The use of cells' biological tools for research purposes allowed, and continues to allow, scientists to discover information about the genomes of living organisms, including humans.

58. The genomes of individual members of a species can differ from each other. The most common form of such genetic variation is a "single nucleotide polymorphism" ("SNP"), wherein a single nucleotide is replaced with a different nucleotide, and the resulting genetic variant is observed in a significant fraction of a population (typically at least 1%). In cancer, the genomes of individual cells that comprise a tumor differ from those of non-cancerous cells. These differences can

take the form of "point mutations" at a single nucleotide position. Sometimes certain genetic sequences are erroneously repeated, deleted, or transposed within a gene, or portions of two separate genes may be spliced together to create a "gene fusion." Some variants in the sequence of a gene can disrupt the protein synthesis process, thereby potentially disrupting the functionality of an organism's cells and the organism itself.

59. DNA sequencing information (including aberrations or "mutations" in the sequence) can be used to "define characteristics of genomes, and delineate differences between them, which, in turn, help us to understand genotype/phenotype relationships," that is, the relationship between an organism's genetic information and its expressed traits. EX1009 (Bronner) at 0001. DNA sequencing therefore has cross-disciplinary applications across a breadth of fields, ranging from genetics to molecular biology, pathology, archeology, forensic sciences, and more.

G. RNA Sequencing

60. Because RNA serves the critical biological function of translating genetic information encoded in DNA into proteins, researchers also have been interested in sequencing RNA. RNA sequencing provides insight into an organisms transcriptome, including, among other things, gene expression and any changes in gene expression due to mutations in DNA.

61. To sequence RNA, scientists use reverse transcriptases to convert RNA

into cDNA, which is more stable and better suited for sequencing. The sequencing methods described below can also be applied to cDNA and thus can be used to elucidate information held by RNA.

H. "First Generation" Sanger Sequencing

62. The first widely adopted method of DNA sequencing was developed by Frederick Sanger in the 1970s. This method, for which Dr. Sanger was awarded the Nobel Prize in 1980, was the dominant approach to DNA sequencing from the late 1970s through the mid-2000s. While the field was disrupted by the arrival of "Next Generation Sequencing" ("NGS") technologies in the mid-2000s, Sanger sequencing is still widely used for certain applications today.

63. Sanger sequencing is a "sequencing by synthesis" ("SBS") technique, meaning it determines the sequence of a single-stranded DNA template by "synthesizing" the template's complementary strand, base by base. Sanger sequencing involves: (1) single-stranded DNA templates, (2) DNA primers, (3) DNA polymerases, (4) "normal" nucleotides (dNTPs) corresponding to each base (G, C, A, T), and (5) nucleotides modified to terminate DNA strand elongation (di-deoxynucleotide triphosphates "ddNTPs") corresponding to each base. The ddNTPs are radioactively- or fluorescently-labeled to allow for detection.

64. Early implementations of Sanger sequencing worked by generating copies of the original DNA template using cloning vectors and *E. coli* and then

separating the resulting template copies (each theoretically having the same sequence) into four separate reactions,[1] one corresponding to each of the four types of ddNTPs (G, C, A, T). In each reaction, a mixture of all four dNTPs and, at a much lower concentration, labeled ddNTPs corresponding to one of the four bases, are mixed with the DNA templates, primers, and and polymerase so that primers anneal to the DNA templates and are extended by the random incorporation of dNTPs and ddNTPs. When a ddNTP is added to a growing primer, synthesis of that strand terminates. Once the four reactions are complete, the results of the reactions are separated by size using gel electrophoresis, with a single lane for each of the four reactions. DNA bands that correspond to DNA fragments with differing lengths are visualized, allowing for the order of nucleotides to be determined based on the length at which the various labelled ddNTPs appear. This process is illustrated in the figure below, which I have reproduced from EX1068 (Gauthier) at 0020.

[1] This was true for early implementations that relied on radioactivity. Once fluorescently labelled ddNTPs could be used, only one reaction was necessary because incorporation of any of the four nucleotides was distinguishable by the color of the labeling fluorophore.

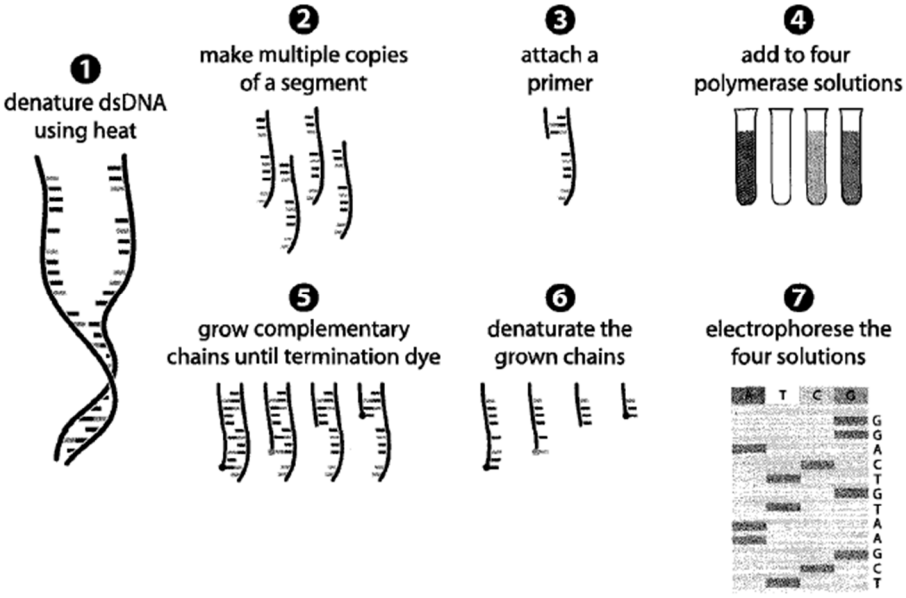


FIGURE 1.2 The Sanger sequencing method in 7 steps. **(1)** The dsDNA fragment is denatured into two ssDNA fragments. **(2)** A fragment of ssDNA is multiplied into millions of copies. **(3)** A primer that corresponds to one end of the fragment is attached. **(4)** The fragments are added to four polymerase solutions. Each solution contains the four types of bases but only one type of termination nucleotide. **(5)** The chain grows until a termination nucleotide is randomly added. **(6)** The resulting dsDNA fragments are denatured to obtain a series of ssDNA of various lengths. **(7)** The fragments are separated by electrophoresis and the sequence is read.

65. Over time, various improvements were made to Sanger sequencing to increase its efficiency, including replacing the ddNTP radioactive labels with fluorescent dyes and replacing gel electrophoresis with automated capillary electrophoresis. EX1050 (Moorthie) at 0001; EX1009 (Bronner) at 0002. This automated Sanger sequencing was effective, but still ultimately limited in throughput, "as sequencing [was] performed on single isolated templates, which means that large-scale sequencing projects [were] expensive and laborious, requiring ligation of target DNAs into cloning vectors, and amplification in *E. coli*." EX1009

(Bronner) at 0001. The throughput of Sanger sequencing methods was also practically constrained by the physical size of the capillary machine required to interpret large numbers of samples. EX1009 (Bronner) at 0001.

66. As a result, as of the early 2000s, while Sanger sequencing remained the primary method for sequencing, there continued to be a significant cost barrier to DNA sequencing at a large scale, as the technique was "too laborious and expensive for routine sequencing of anything more than a few genes." EX1050 (Moorthie) at 0001.

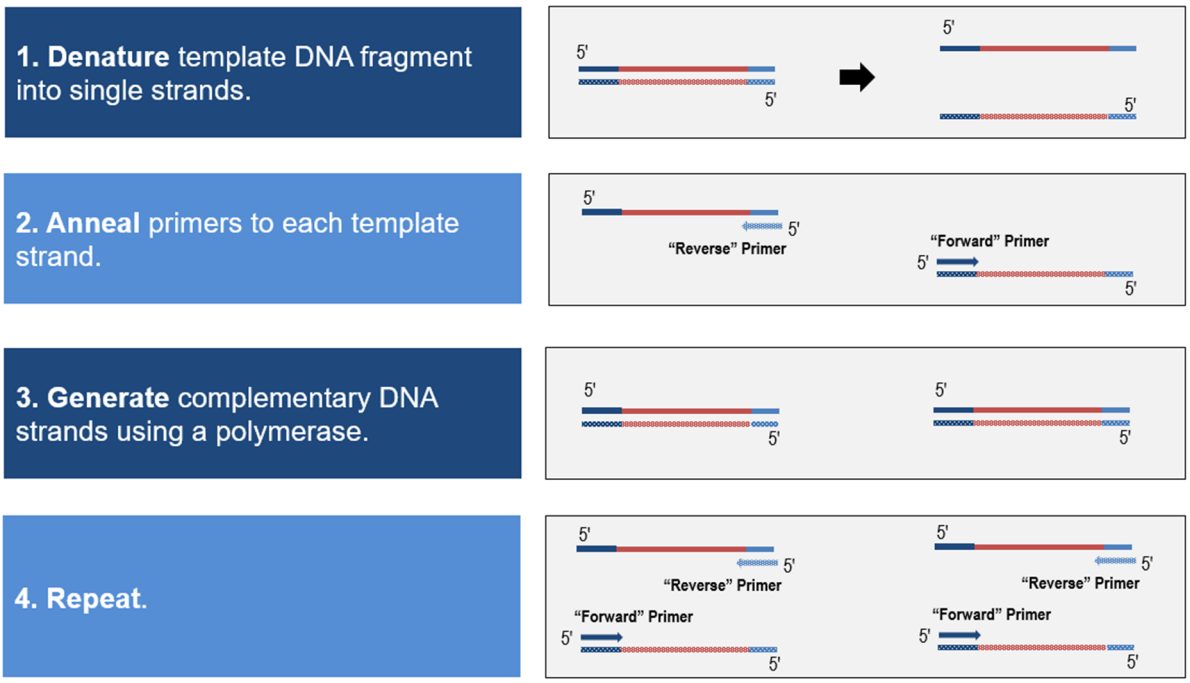
I. Polymerase Chain Reaction ("PCR")

67. In the mid-1980s, Kary Mullis developed a technique known as a "Polymerase Chain Reaction" or "PCR" that allowed for the exponential copying of DNA. This technique, for which Dr. Mullis was awarded the Nobel Prize in Chemistry in 1993, enabled researchers to selectively target and exponentially amplify target nucleic acid fragments, which was a groundbreaking advancement with a wide range of research applications, from genotyping to gene expression analysis.

68. PCR involves temperature-dependent steps, in which free nucleotides in solution, "forward" and "reverse" oligonucleotide primers, and a DNA polymerase are used to exponentially generate copies of double-stranded nucleic acid fragments. The primers are added to the reaction at an excess relative to the

original "template" fragments.

69. The below simplified figure depicts the PCR process as applied to a single double-stranded DNA fragment. DNA strands are represented by the colored bars. Each colored bar has three colors, representing, from left to right, an adaptor containing a primer binding site, a DNA fragment, and a second adaptor containing a second primer binding site. While the figure below depicts only one DNA fragment, this process typically involves a single reaction containing thousands or millions of double-stranded DNA fragments undergoing the same chemical steps in parallel. As demonstrated by the below figure, the PCR amplification process is exponential in nature, in that the number of double-stranded fragments in the reaction doubles with each cycle.



70. The PCR reaction requires "thermocycling" because the denaturing, annealing, and DNA synthesis chemical processes occur under specific temperature conditions. For example, heat is required for DNA strands to denature. Then, heat is reduced to allow oligonucleotide primers to anneal to the denatured strands. Finally, the heat is increased slightly to allow for the DNA polymerase to generate complementary strands starting from the primers. The entire process is then repeated by heating again to denature the newly formed double-stranded products.

71. PCR can be used to selectively target and amplify a particular region or gene of interest by using primers specific to known sequences "upstream" and "downstream" of the desired target region.

72. PCR can also involve ligating "adaptors," that is, synthetic oligonucleotides of a known sequence, as I discussed above in Section IV.E, to both ends of each fragment (as depicted in the shades of blue in the figure above). Primers with sequences complementary to the known adaptor sequences are then used in the PCR reaction to anneal to and thereby amplify only those fragments containing adaptors.

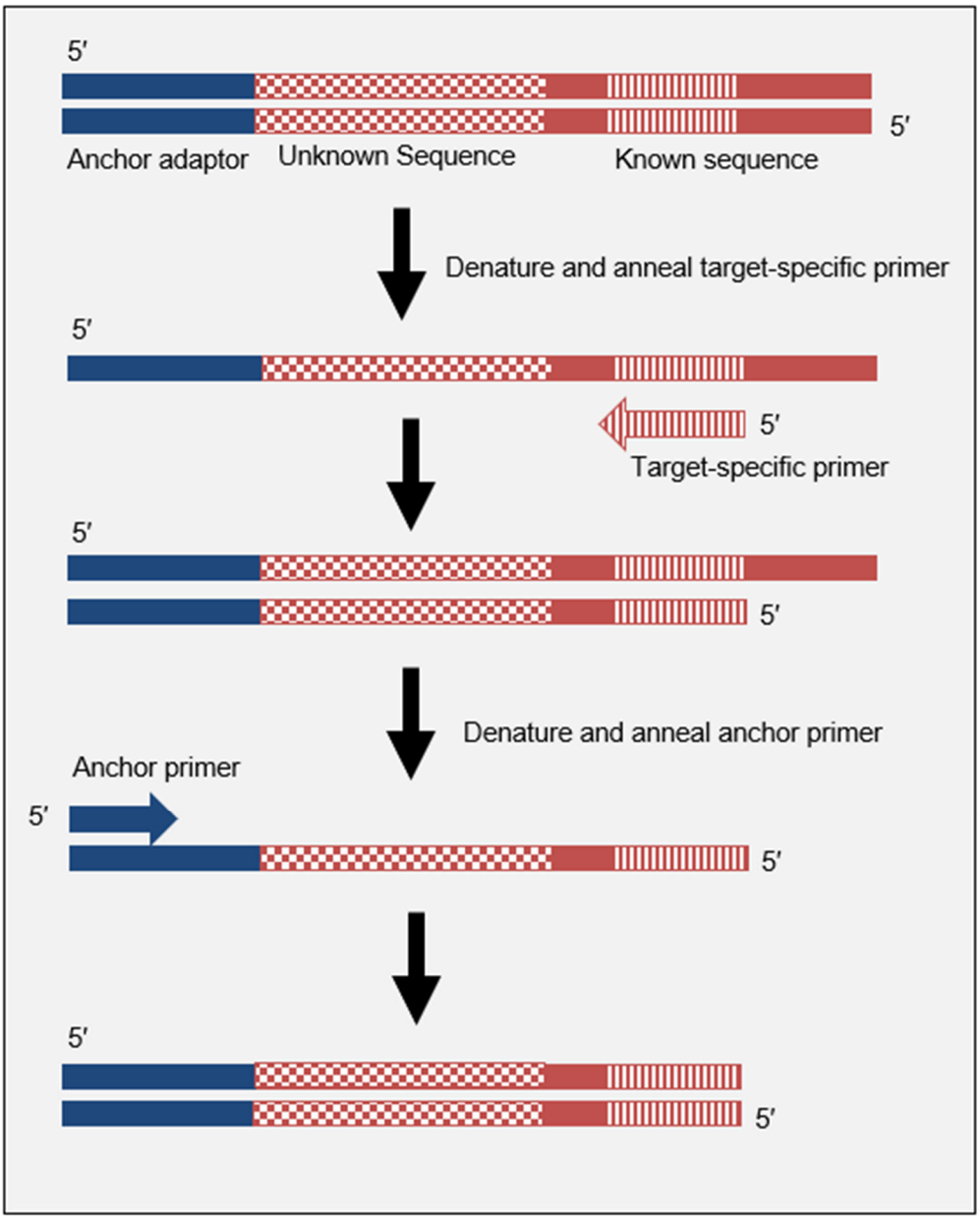
J. "Anchored PCR" and "Genome Walking"

73. By the early 1990s, researchers had begun modifying and adapting the PCR technique to new applications. One application was to amplify regions of DNA or RNA containing both a known sequence and an adjacent unknown sequence.

Using traditional targeted PCR, both a target-specific "forward" and target-specific "reverse" primer are required, and thus the beginning and the end of the gene of interest must be known. Scientists, however, were interested in a method of enriching regions where only the beginning or the end of the sequence of interest was known while other portions of the sequence were unknown or variable. Determining the sequence of those unknown or variable regions can help explain certain biological processes or phenomenon, such as cancer. But scientists first had to overcome the challenge of designing a primer for a sequence that was unknown or variable.

74. Scientists thus developed PCR-based methods for selectively targeting unknown regions adjacent to a known sequence using a technique known as "single-sided PCR," "ligation-anchored PCR" or "anchored PCR." EX1058 (Troutt) at 0001. As shown in the figure below, an anchored PCR reaction involves two primers: (1) the "anchor primer," (dark blue arrow below) which is complementary to a fixed "anchor" sequence, often in an adaptor (dark blue segment below) ligated to the nucleic acid fragment (red below), and (2) the "target-specific primer," (striped red arrow below) which is complementary to a known sequence of interest (striped red segment below) in the fragment. The result of this PCR reaction is to amplify fragments with the known sequence of interest on one side and the anchor adaptor on the other, with potentially an unknown sequence (red checkered segment

below) between the two.



75. For example, in 1992, Troutt *et al*. published a ligation-anchored PCR method for selectively enriching an unknown sequence adjacent to a known sequence involving three steps: (1) synthesizing cDNA from mRNA, (2) ligating an adaptor with a known artificial sequence to the 3′ end of the cDNA strand, and

(3) PCR amplifying the adaptor-ligated product using one primer with sequence complementary to the adaptor and one primer with a known sequence contained in the cDNA. EX1058 (Troutt) at 0002. The result of this process is the creation of a selectively enriched sample of amplified DNA containing both the known fragment sequence and the unknown sequence directly adjacent to the known sequence. This process is illustrated in the below figure, which I have reproduced from EX1058 (Troutt) at 0002, in which the "known" region is denoted by a hatched box, and the "unknown" sequence of interest is directly 5′ of that region on the original mRNA strand (Step A in the figure below), and 3′ of that region on the generated cDNA strand (Step B in the figure below):

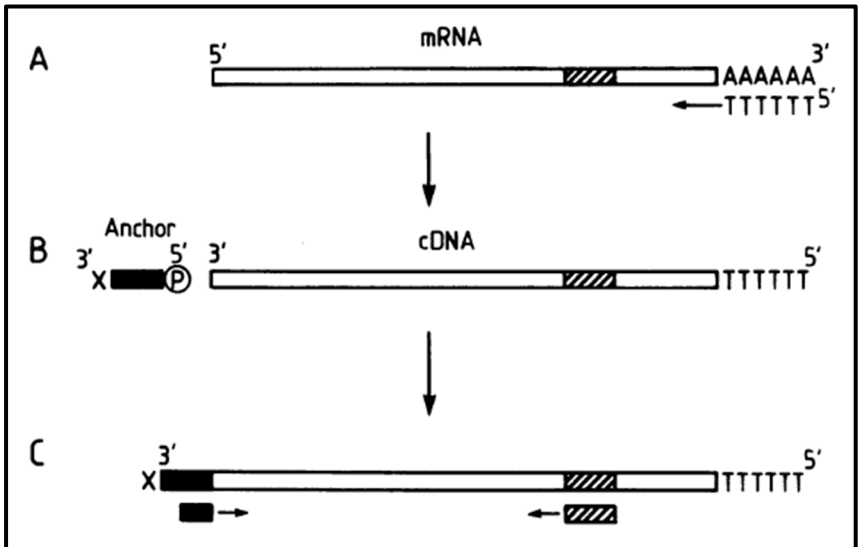
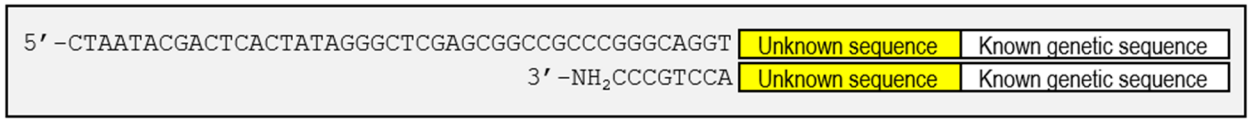


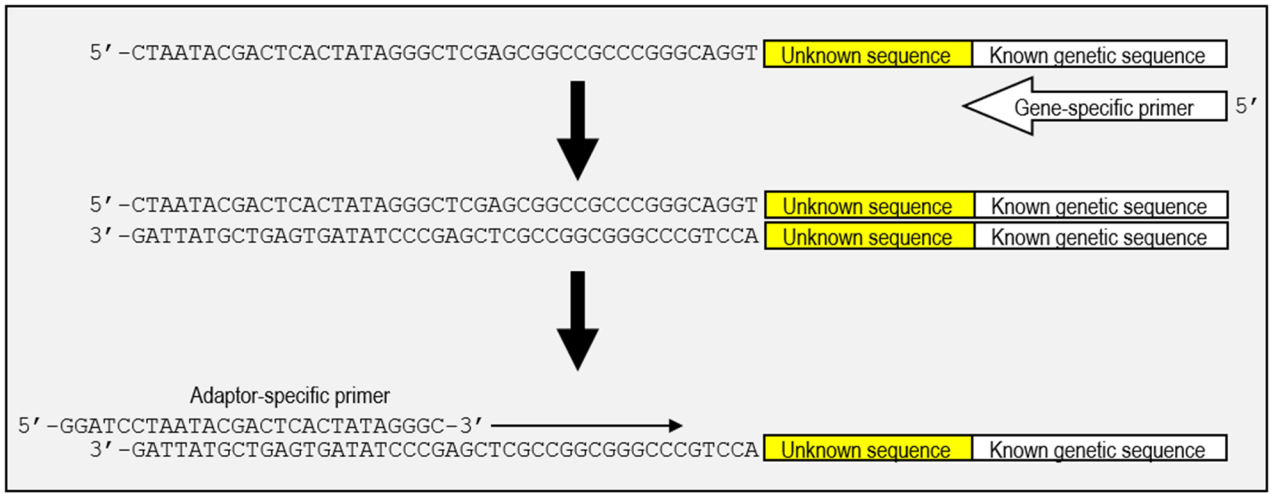
FIG. 1. Protocol for LA-PCR. (*A*) First-strand cDNA synthesis was performed with avian myeloblastosis virus reverse transcriptase and oligo(dT) as primer. (*B*) The 5′-phosphorylated, 3′-end blocked anchor oligonucleotide was ligated to first-strand cDNA with T4 RNA ligase. (*C*) PCR amplification was carried out with oligonucleotide primers specific for a known sequence within the cDNA (hatched box) and for a sequence within the anchor (solid box).

76. By the mid-1990s, "anchored PCR" methods had been adapted for use with genomic DNA. This enabled researchers to isolate and detect, for example, the sequences of unknown or variable genetic mutations that were known to commonly occur at a particular location within a gene. For example, in 1995, Siebert published a method for a technique known as genome "walking." EX1007 (Siebert) at 0001-02. This technique enabled "walking from a known region to an unknown region" in genomic DNA by selectively targeting and amplifying a "known" sequence of interest in order to also enrich an unknown adjacent sequence. EX1007 (Siebert) at 0001.

77. In the Siebert method, genomic DNA is "digested," or cut into fragments, using a type of enzyme known as a "restriction enzyme." EX1007 (Siebert) at 0001. Restriction enzymes are capable of cleaving DNA at sequence-specific sites. As part of this cleavage process, the enzyme leaves the phosphate on the "new" 5′ ends intact, so that ligation to the "new" 5′ ends is possible. Then double stranded adaptors are ligated to the resulting fragments of double stranded genomic DNA using a DNA ligase. EX1007 (Siebert) at 0002. Siebert used an adaptor wherein the antisense strand was shorter than the sense strand, and blocked at the 3′ end to prevent extension of the short adaptor strand, EX1007 (Sibert) at 0001, which I have depicted in the figure below.



78. The authors then introduced a "gene specific primer" (white arrow below) complementary to a known sequence in the genomic DNA (white box below) and a primer with the sequence of the longer 5′-ligated adaptor. EX1007 (Siebert) at 0001. The authors then used PCR with those two primers to selectively amplify the unknown sequence between the adaptor on one end and the known genetic sequence on the other end of each fragment, EX1007 (Siebert) at 0001-02, which I have depicted in the figure below.



79. The result of the process was a selectively enriched library of fragments containing both the known genetic sequence and the adjacent unknown sequence. EX1007 (Siebert) at 0002.

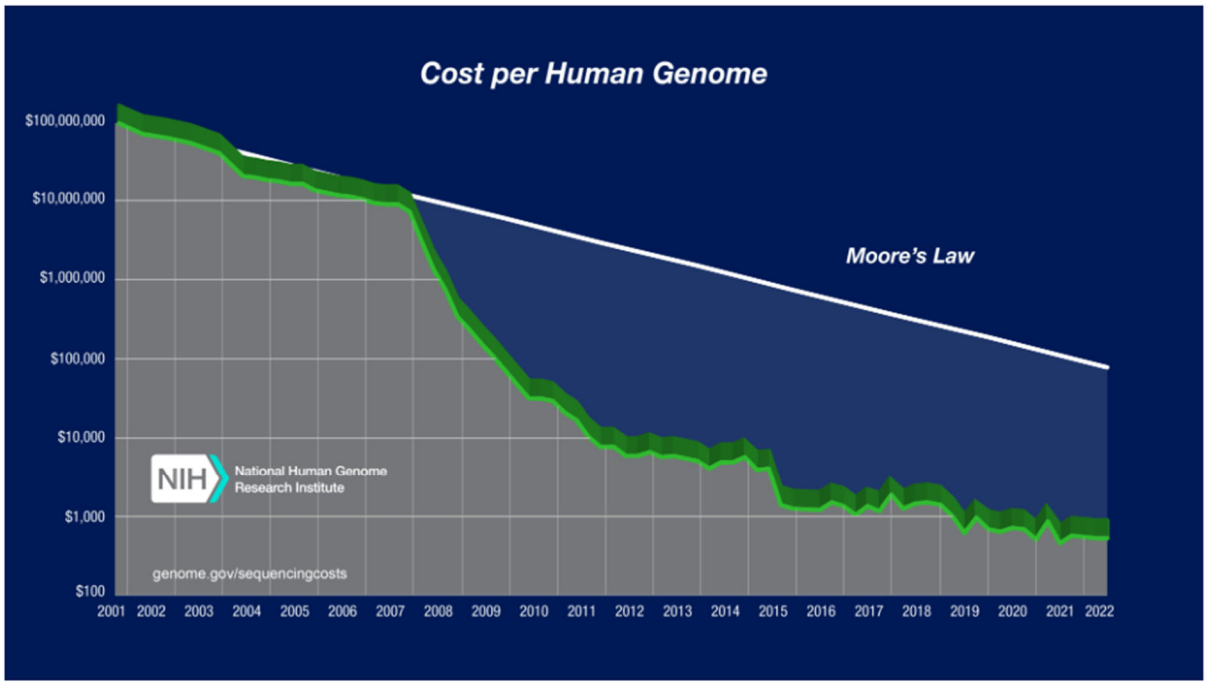
K. "Next Generation Sequencing"

80. Beginning in the mid-2000s, sequencing technology was revolutionized

by the introduction of "next generation sequencing" ("NGS"), which drastically reduced the time and cost associated with DNA sequencing. EX1050 (Moorthie) at 0001-02. While each Sanger sequencing "run" was limited to sequencing copies of a single, predefined target (up to around 1,000 bases), NGS, also referred to as "massively parallel," "high throughput," "deep," or "clonal" sequencing, enabled an "explosion in capacity" (also known as "throughput") by introducing the ability to sequence billions of fragments originating from millions of distinct targets in a single sequencing run. EX1050 (Moorthie) at 0002; EX1008 (Caruccio) at 0001.

81. NGS greatly decreased the cost of sequencing and opened the door to many new applications of sequencing that would previously have been cost prohibitive. EX1050 (Moorthie) at 0001-02. The Human Genome Project is illustrative of the cost reduction from Sanger sequencing to NGS. The Human Genome Project was completed using primarily Sanger sequencing and required hundreds of research groups working together over more than a decade, costing hundreds of millions of dollars. EX1009 (Bronner) at 0001. The subsequent mid-2000s arrival of NGS methods that allowed for the sequencing of massive amounts of DNA all at once removed those cost barriers, suddenly making DNA sequencing of whole genomes an accessible—even routine—technology for most academic research labs. The Human Genome Project endeavored to sequence and assemble the human genome from scratch, producing a "reference genome" that is used by

researchers to this day. By 2011, following the introduction of NGS, the cost of re-sequencing the human genome for mapping to the reference produced by the Human Genome Project had dropped by orders of magnitude, as shown below in the figure I have reproduced from EX1071 (NIH DNA Sequencing Costs Data) at 0002.



Cost per genome data - 2022

82. Multiple NGS techniques all launched in the early 2000s. These techniques differed slightly in terms of the precise mechanism of achieving massively parallel sequencing, but all shared certain common features, specifically a reliance on a solid support to fix fragments in a spatially defined location and the use of modern optical technology to image or measure signals from isolated groups of fragments at much higher resolution than was previously possible.

83. Roche/454 pyrosequencing ("454") was introduced in 2004 and

became the first commercially available NGS platform in 2005. EX1070 (Margulies) at 0001-06; EX1009 (Bronner) at 0002; EX1014 (Masoudi-Nejad) at 0022. By 2011, a handful of additional NGS "platforms" had come to market, each with slightly different chemistry and machines, including Solexa/Illumina reversible terminator sequencing ("Illumina") and Life Technologies/Ion Torrent semiconductor sequencing ("Ion Torrent"), all of which I describe below in additional detail.

84. Despite variations in each platform's precise chemistry, the POSA would have understood that all three of the aforementioned NGS platforms operated according to the same high-level process:

- *Library Preparation*. The 454, Illumina, and Ion Torrent NGS platforms all begin with a library of nucleic acid fragments, typically a few hundred bases in length. These fragments are attached to platform-specific adaptors with synthetic fixed, or "universal," sequences that are then used for later chemistry steps and to anneal the fragments to probes fixed to a solid support. EX1009 (Bronner) at 0003.
- *Annealing to a Surface*. The 454, Illumina, and Ion Torrent NGS platforms all then anneal single-stranded adaptor-ligated fragments to oligonucleotide "probes" fixed to a solid surface, such as a chip or a bead. The probes have sequences complementary to the adaptor sequences. The

surface contains millions or billions of identifiable locations, such as a single bead or position on a chip. The target fragments are hybridized to the surface such that each spatially identifiable location is ideally initially populated by, at most, a single, distinct target molecule, for example a single molecule is hybridized to a single bead or a single identifiable position on a chip. EX1050 (Moorthie) at 0002.

- *On-Surface Amplification*. The 454, Illumina, and Ion Torrent NGS platforms all then amplify the adaptor-ligated fragments that have been localized on the surface (for example, using PCR), creating a "cluster" of copies of the original fragment, for example on a single bead or at a certain location on a chip. This cluster amplification creates a group of enough fragments to generate a detectable signal during sequencing.
- *Massively Parallel Sequencing*. The 454, Illumina, and Ion Torrent NGS platforms all then sequence the clusters of localized amplified fragments using a sequencing-by-synthesis process. Similar to Sanger sequencing, a subset of NGS platforms use the action of DNA polymerase to add nucleotides complementary to the fixed fragments. Other NGS platforms use DNA ligase to add oligonucleotides complementary to the fixed fragments. But unlike Sanger sequencing, NGS platforms are able to detect each individual nucleotide-incorporation event at each individually

positionally-defined cluster as it occurs. Most NGS platforms leverage modern computer and imaging technology to measure each cluster individually in a single sequencing "run," for example by using a computer-assisted microscope to optically detect light emitted by the incorporation of modified nucleotides at a particular cluster of fragments. As I explain further below, some NGS methods like the Ion Torrent platform use microfabricated electronics to detect small changes in pH that occur after incorporation of a nucleotide by DNA polymerase, rather than optical methods.

- *Computational Processing and Sequence Determination*. The 454, Illumina, and Ion Torrent NGS platforms also all take advantage of modern computer technology to interpret the billions of nucleotide incorporation events at the various positions to output a data file of raw sequencing data (comprising sequencing "read" data for each individual cluster) for the entire library of nucleic acid fragments and use computational analysis to "align" the individual reads to a reference (*e.g.*, whole genome or whole transcriptome) sequence. One common file format for raw sequencing data, including for Illumina devices, is FASTQ. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034-35.

85. The ability of NGS platforms to fix fragments to much smaller spatial

positions (for example, to beads or a grid on a chip) and then visualize the resulting spatially-segregated clusters during simultaneous SBS reactions represented a breakthrough over Sanger sequencing's reliance on capillary-based methods. Because NGS platforms permitted orders of magnitude more fragments to be sequenced at the same time in much less physical space and with significantly less reagent volume per reaction, NGS platforms had drastically higher throughput in terms of the number of fragments sequenced in a single run and lower costs per fragment sequenced.

1. NGS Platforms

86. A major component of my PhD and postdoctoral research at Harvard University between 2007 and 2011 was on the development of an NGS technology called fluorogenic pyrosequencing. Sims, P.A., Greenleaf, W.J., Duan, H., Xie, X.S., *Fluorogenic DNA Sequencing in PDMS Microreactors*, NATURE METHODS, 8, 575–580 (2011). As a trainee, I also gained experience with commercial NGS platforms starting in 2009 using the 454 Genome Sequencer FLX and later using the Life Technologies SOLiD sequencer and the Illumina Genome Analyzer and HiSeq 2000. The two instruments from 454 and Life Technologies were purchased by the laboratory in which I was performing my PhD in postdoctoral studies, and the Illumina sequencers were available in a core facility at Harvard. When I started my own research laboratory at Columbia University in 2012, I used exclusively Illumina

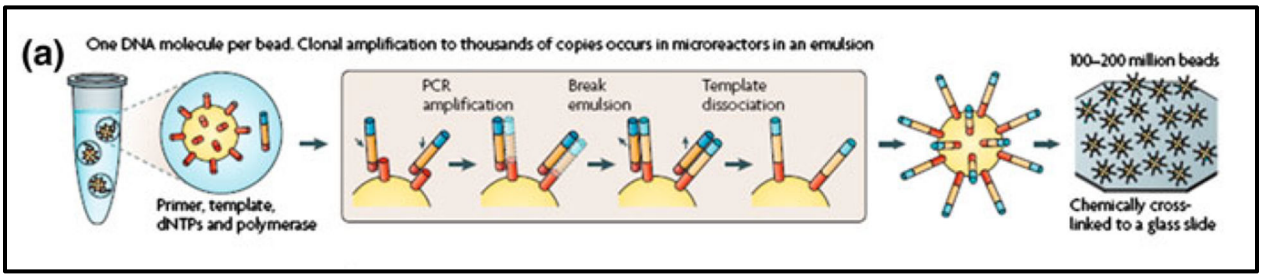
sequencers available in the Sulzberger Columbia Genome Center (Illumina HiSeq 2000, HiSeq 2500, HiSeq 4000, NovaSeq 6000, NextSeq 500, and MiSeq) until 2023 when we began additionally working with a new platform from Element Biosciences. Throughout this declaration, I rely on my own personal knowledge of the functionality of NGS platforms based on my own experience. Where useful, I also include information and figures from Masoudi-Nejad's "Next Generation Sequencing and Sequence Assembly" textbook. EX1014. Although Masoudi-Nejad's textbook was published in 2013, it accurately reflects information a POSA would have known about NGS platforms and techniques as of 2011.

a. Roche/454 Pyrosequencing

87. On the 454 platform, single-stranded adaptor-ligated fragments from a serially diluted solution are hybridized to probes fixed to the surface of hundreds of millions of beads, such that each bead is ideally initially populated by, at most, a single fragment. Each fragment anneals at its 3′ end to the probe, so that, as part of the emulsion PCR step described below, the probe can act as a primer and be extended (from 5′ to 3′) to generate a fragment complementary to the hybridized fragment.

88. Each bead is then encapsulated in a droplet formed by a water/oil emulsion that effectively turns each bead into its own micro-reactor containing the required reagents for PCR. The fragments on each bead are amplified in an

"emulsion PCR" reaction, resulting in the surface of each bead being covered in copies/amplicons of the original fragment, as shown below in the figure I have reproduced from EX1014 (Masoudi-Nejad) at 0022.



89. This emulsion PCR process was much more cumbersome and inconvenient than the "set it and forget it" bridge amplification process employed by Illumina that I describe below in Section IV.K.1.b. I used a 454 Genome Sequencer during my training at Harvard in 2009, which relied on emulsion PCR. This procedure required significant hands-on time compared to the bridge amplification process used by Illumina sequencers. In one of my later projects as a trainee in 2011, I switched to using an Illumina HiSeq 2000, which was available in a core facility at Harvard. By this time, the Illumina sequencers were not only easier to use, but also had significantly higher throughput than the 454 Genome Sequencer. While the 454 Genome Sequencer provided longer read lengths than the Illumina sequencers at the time, this advantage was outweighed by the ease-of-use and throughput available from Illumina platforms for many applications.

90. Following emulsion PCR, the emulsion is broken on each bead and the

beads containing the amplified clusters are rolled across a surface containing millions of wells such that each well is occupied by a single bead. The clusters on the beads then undergo a "pyrosequencing" SBS reaction in which the incorporation of each individual nucleotide is measured by optically detecting light emitted as part of enzymatic activity associated with the release of a pyrophosphate during each nucleotide incorporation event. EX1014 (Masoudi-Nejad) at 0021-22. The emitted light is generated only temporarily, such that, unlike on Illumina platforms, which use fluorescent dNTPs that fluoresce until they are removed, all fragments on a 454 machine must be imaged in "real time," *i.e.*, in a single imaging frame, in order to capture the emission of light associated with each incorporation event. This limitation presented a technological constraint on the potential throughput (number of fragments that could be sequenced at once) of the 454 technology.

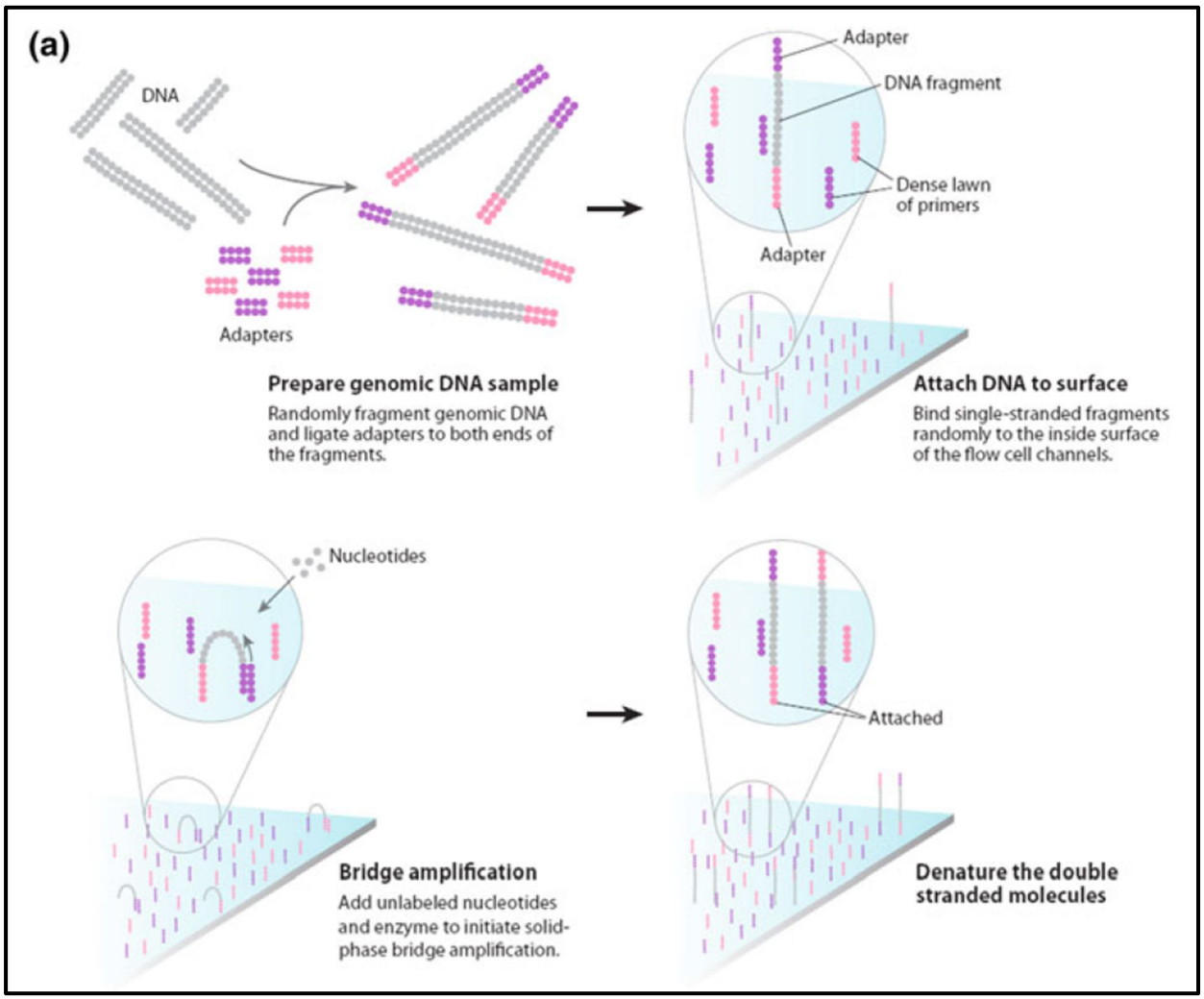
91. As of 2011, a further limitation of the 454 platform was its inability to accurately sequence long homopolymer stretches (that is, strings of a single nucleotide type). In homopolymer stretches, the 454 machine detects a signal intensity much higher than that of a single nucleotide. This higher intensity signal can only be precisely quantified for fewer than 10 consecutive nucleotides. EX1014 (Masoudi-Nejad) at 0021.

b. Solexa/Illumina Reversible Terminator Sequencing

92. On the Illumina platform, template DNA molecules are hybridized to a

dense "lawn" of "universal" probes fixed to a glass surface known as a "flow cell." EX1050 (Moorthie) at 0002. Both amplification and sequencing take place on the flow cell. EX1050 (Moorthie) at 0002. There are two different types of immobilized probes (a "P5" and "P7") probe, with each having a sequence complementary to one of the two distinct adaptors ligated to each end of the template DNA molecules. Therefore, on the Illumina platforms, both the sense and antisense strands of the double-stranded templates are captured on the flow cell at their 3′ end in their single-stranded form.

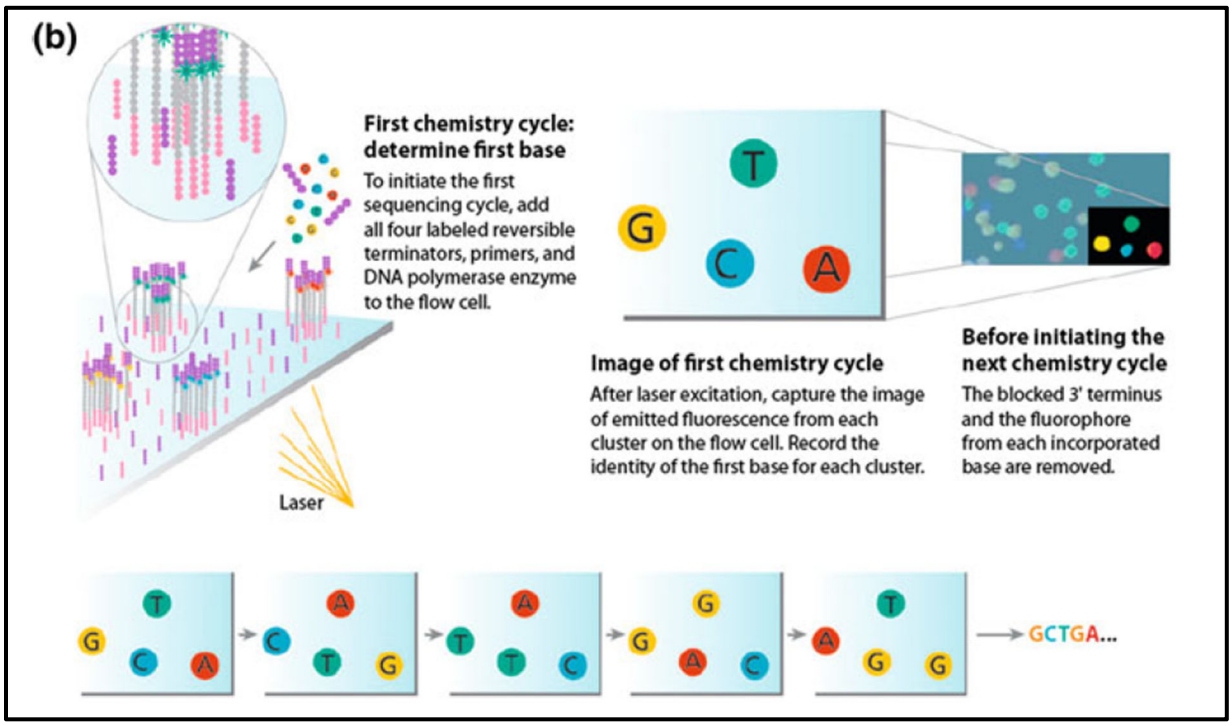
93. Clonal amplification to create "clusters" of copies of each individual fragment is then performed via a "bridge amplification" process wherein the universal probes fixed to the surface are used as forward and reverse PCR primers, EX1050 (Moorthie) at 0002, as shown below in the figure which I have reproduced from EX1014 (Masoudi-Nejad) at 0024.



94. This bridge amplification technique was a significant innovation, because it was much more hands-off than the emulsion PCR amplification process I describe above, which was used by the 454 and Life Technologies/Ion Torrent semiconductor sequencing platforms.

95. The bridge amplification process results in "clusters" of identical single-stranded DNA molecules that are ready for sequencing. Sequencing on the Illumina platforms is performed via SBS using fluorescently labelled "reversible

terminator nucleotides." EX1050 (Moorthie) at 0002. Sequencing primers are annealed to the single-stranded template molecules at their 3′ end (the end distal from the surface of the flow cell) and a polymerase is used to incorporate fluorescently-labeled "reversibly blocked" nucleotides one at a time. EX1014 (Masoudi-Nejad) at 0023. The nucleotides are temporarily blocked at their 3′ termini so that only one nucleotide can be added to each fragment at a time. EX1014 (Masoudi-Nejad) at 0023. As each nucleotide is incorporated, it fluoresces a certain color unique to each of the four bases. EX1014 (Masoudi-Nejad) at 0024. A scanning optical image is taken after each incorporation event of all of the fragments on the entire flow cell to determine which base was added in each cluster. EX1014 (Masoudi-Nejad) at 0024. After image capture, the blocked 3′ terminus and fluorescent dye on each nucleotide are chemically removed, allowing the next round of labeled reversibly blocked nucleotides to be added to the chain. This process is illustrated in the image below, which I have reproduced from EX1014 (Masoudi-Nejad) at 0024.



96. The Illumina sequencing machine then optically detects the fluorescence of each newly incorporated nucleotide, but because the fluorescence associated with each nucleotide incorporation event lasts longer than the corresponding ephemeral luminescent signal emitted on the 454 platform, not all fluorescent signals have to be captured in "real time" in a single frame. Instead, Illumina uses a "scanning" optical technology to detect the fluorescent signals. As a result, unlike for 454, the throughput of the Illumina platform is not constrained by the physical size of the optical imager, *i.e.*, the number of pixels that can be captured in a single frame. This allowed Illumina to quickly outpace 454 in terms of throughput.

97. Illumina devices output raw sequencing data in particular file formats,

including FASTQ. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35. These files contain the sequencing data for each read performed during a sequencing run. EX1061 (Kircher Thesis) at 0080–81; EX1062 (Smith) at 0034–35. When several samples are sequenced together, a process I discuss below in Section IV.K.2.a, the Illumina device recognizes an artificial sequence called a "sample index" and outputs separate FASTQ files for each sample, a process sometimes called "demultiplexing." EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35. Sequencing reads are demultiplexed and FASTQ files generated before further data analysis takes place, EX1062 (Smith) at 0034–38; EX1063 (Won) at 0003–04, including identification and removal of duplicate sequencing reads, EX1054 (Loeb) at 0012–15.

c. Life Technologies/Ion Torrent Semiconductor Sequencing

98. The semiconductor platforms use a similar emulsion PCR bead-based technique as Roche/454 for capturing and amplifying samples.

99. For the actual sequencing, semiconductor platforms use an SBS chemistry that does not involve optical imaging. Semiconductor platforms use a microchip filled with a high-density array of micro-machined wells as a surface. Beneath the wells is an "ion-sensitive" layer, and beneath that layer is an "ion sensor." Each well contains a "cluster" of amplicons of a single target molecule.

The microchip surface enables the sequencing machine to detect the change in pH in each individual well caused by the release of a hydrogen ion following the incorporation of each nucleotide onto a strand of DNA. EX1014 (Masoudi-Nejad) at 0027-28.

2. Library Preparation for NGS

100. All NGS platforms require an initial step in which the nucleotides that are being prepared for sequencing are fragmented and attached to platform-specific "universal" adaptors with fixed sequences. In 2011, NGS sequencing platforms followed a "common three-step workflow" for library preparation. EX1008 (Caruccio) at 0001-02. This workflow typically included: (1) fragmentation of the original sample nucleic acids, (2) end-repair of the fragments "to have blunt ends or A tails," and (3) the addition of "platform-specific adaptors" to the ends of the fragments. EX1008 (Caruccio) at 0001-2; EX1009 (Bronner) at 0003.

101. In many instances, library preparation also included PCR amplification prior to annealing the fragments to the NGS platform (where they would undergo an additional "cluster amplification" step, as I described above).

102. One purpose of the adaptor sequences appended to the nucleic acid fragments was to enable the initial PCR amplification steps. Another purpose of the adaptors was to allow the fragments to anneal to the probe sequences on the solid surface of the NGS machine, such as the P5 and P7 sequences I discussed above in

relation to Illumina platforms. In addition, adaptors contained a synthetic sequence complementary to the sequencing primer that would be used in the SBS reaction.

103. Adaptors also frequently contained other synthetic sequences that would be used for other purposes. As of 2011, two examples of additional adaptor sequences were (1) sequences for sample indexing, which allowed fragments from multiple original samples to be pooled together, and (2) sequences that acted as unique identifiers, which were called "molecular tags," among other things, and served to identify PCR copies generated during library preparation that originate from the same DNA fragment in the unamplified sample.

a. Barcodes for Sample Indexing

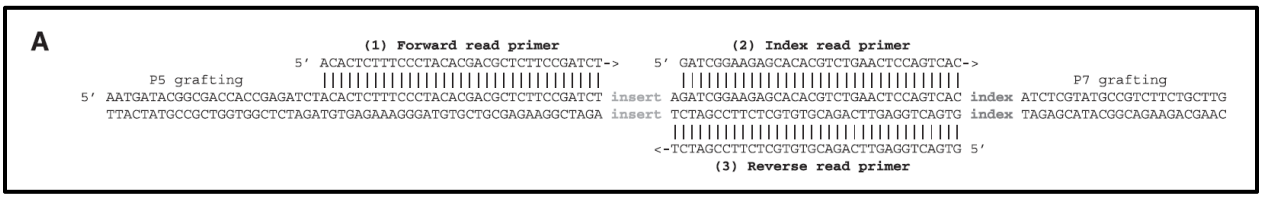
104. Because NGS platforms allowed a very large number of fragments to be sequenced at once, NGS permitted sample "multiplexing"—the pooling and sequencing of multiple samples together—to a much greater extent than was possible with Sanger sequencing. For certain applications, the number of fragments that required sequencing in a particular sample was much smaller than the amount of physical space on the NGS device. In order to efficiently use the space on the device, scientists began pooling multiple samples together prior to sequencing, which required a method of assigning each sequencing read to a particular sample.

105. In 2011, and continuing until today, matching a sequencing read to its sample of origin is generally accomplished using a set of synthetic sequences known

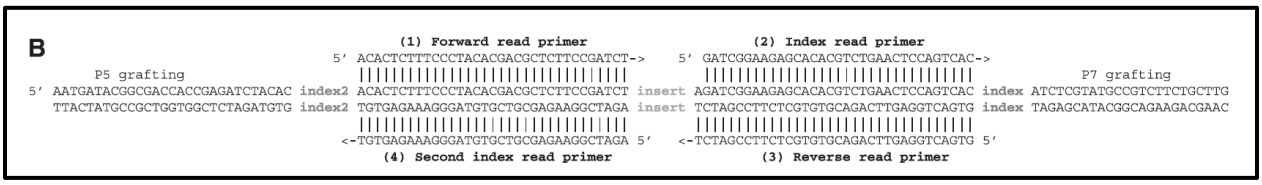
as sample indexes. A sample index is a known nucleotide sequence ranging from a few to tens of bases that is incorporated into the adaptor(s) attached to a particular set of fragments from a particular sample. Each sample index sequence is unique to a particular sample. When the samples are pooled and subsequently sequenced, a particular fragment can be associated with the original sample using the sample index sequence, a process known as "demultiplexing."

106. For example, adapters containing sample index sequence X can be ligated to nucleic acid fragments generated from Sample No. 1 and adapters containing sample index sequence Y can be ligated to nucleic acid fragments generated from Sample No. 2. Sample Nos. 1 and 2 can then be pooled and sequenced together, after which the sequenced samples can be separated based on their sample index sequence, where Sample No. 1 is identified by X and Sample No. 2 is identified by Y.

107. Sample indexes for 454 sequencers were in use at least as of 2007, as discussed in Parameswaran 2007. EX1073 (Parameswaran) at 0001-09. Sample indexes for Illumina sequencers were in use at least as of 2008, as discussed in Craig 2008. EX1072. Initially, sample indexes in Illumina adaptors were incorporated into the P7 adapter, as shown in the figure below, which I have reproduced from EX1055 (Kircher) at 0002.



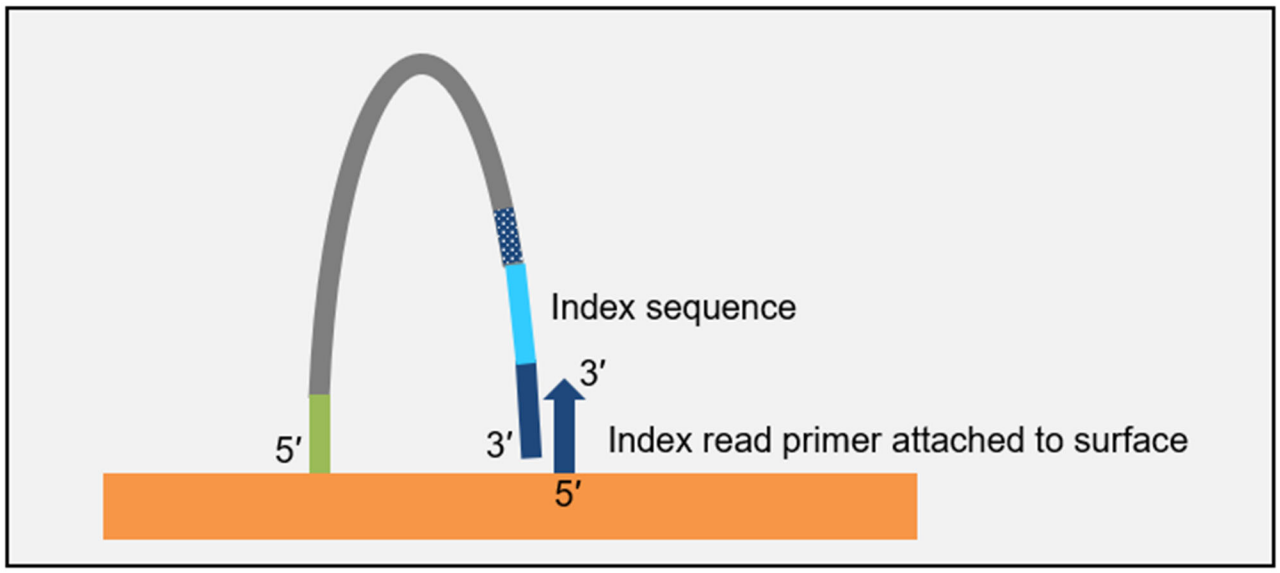
108. But as multiplex sequencing became increasingly dominant, and interest grew in using multiplex sequencing in the context of searching for rare sequence variants, strategies for enhancing the accuracy of multiplex sequencing were developed. EX1055 (Kircher) at 0001. One such strategy relied on adding an additional sample index, which for Illumina sequencing was incorporated into the P5 adapter, as shown in the figure below, which I have reproduced from EX1055 (Kircher) at 0002.



109. The additional sample index meant that each sample was uniquely labeled with two separate short oligonucleotide sequences, allowing additional samples to be pooled together and demultiplexed with high fidelity. Because error rates tend to increase with each nucleotide incorporation or "cycle" of SBS, using two short indexes that could be sequenced in separate reads was advantageous over using a single, long index.

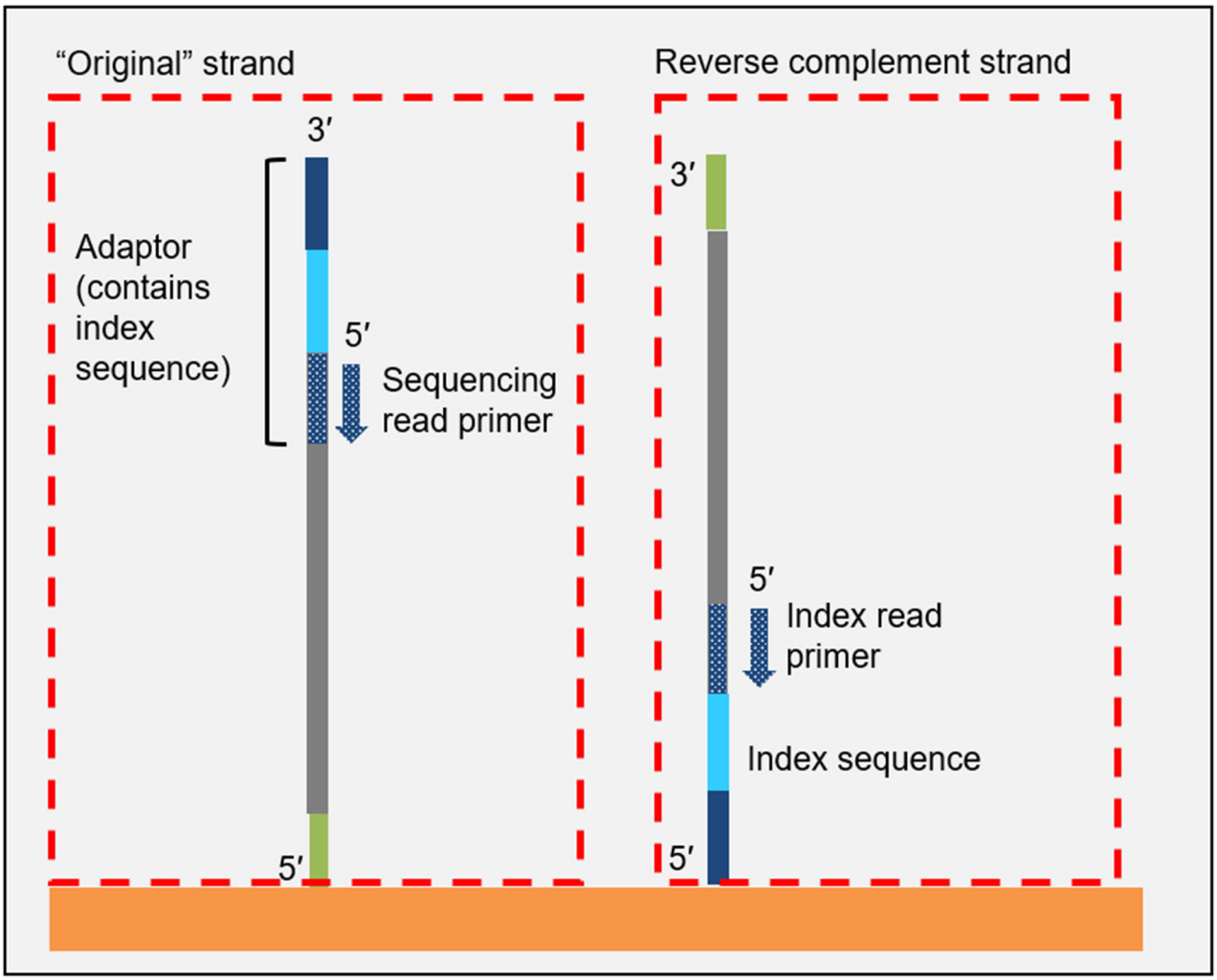
110. The process for sequencing the sample index differed based on the NGS platform used. Within Illumina devices available by 2013, there were several

different workflows for sequencing the sample index. For example, when using an Illumina MiSeq, the sample index is read using an oligonucleotide attached to the surface of the flow cell as the sequencing primer. EX1064 (Kozich) at 0003. The oligonucleotide bound to the flow cell that anneals to the P5 portion of the adaptor is sometimes referred to as the "P5 grafting" sequence, and the bound oligonucleotide that anneals to the P7 portion of the adaptor, the "P7 grafting" sequence. As shown in the figure below, the P5 grafting sequence functions as a sequencing primer and anneals to the adaptor region upstream in the 3′ direction from the sample index (cyan, below) and the nucleic acid fragment (gray, below). The SBS reaction that "reads" the sample index is primed from that position.



111. Other Illumina devices available by 2013, such as the Genome Analyzer II, sequenced the sample index on the reverse complement strand of the nucleic acid fragment. In this workflow, a sequencing primer anneals to a region of

the adaptor sequence located between the fragment (gray) and the index sequence (cyan) and reads the sample index away from the fragment, as shown below. EX1055 (Kircher) at 0002. This workflow also used two different read primers: one for reading the sequence of the nucleic acid fragment (on the "original" strand) and another for reading the sequence of the sample index (on the reverse complement strand).



112. The '399, '357, and '241 patents, which share a specification, refer to

a sample index as an "indexing site," defined as "a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing." EX1003 ('399 Patent) at 4:34–38. (As I discuss below in Section VI.B, certain definitions in the '399, '357, and '241 patents, including "indexing site," are defined the same.)

b. Molecular Tags

113. Shortly following the introduction of NGS, scientists also started adding sequences to adapters to help to identify the PCR copies generated during library preparation that originated from the same DNA fragment in the unamplified sample. Scientists in the field recognized that there were many benefits to being able to ascertain whether a group of one or more sequences created during library preparation were copies of the same original fragment in a sample prior to amplification (e.g. by PCR). The use of random sequence tags or large libraries of pre-defined sequence tags in adaptors allowed scientists to determine whether a particular change observed in a sequence (compared to other fragments or a reference genome) was the result of a mutation in the original fragments or an error introduced during PCR or SBS. In addition, for certain applications, it was useful to be able to count the number of original fragments; for example, to count the number of copies of a particular gene in an original sample.

114. Scientists understood that NGS methods could introduce errors into sequences at several possible points. One possible source of error was library preparation itself. Library preparation for NGS often requires multiple rounds of PCR. PCR can cause multiple different types of errors, including mis-priming events where a primer anneals to a template that is not completely complementary creating a fusion of the incorrect primer to the remainder of the template sequence; point mutations where the polymerase adds an incorrect nucleotide to the growing strand; or "template switching" events where a polymerase starts from one template strand and picks up another template strand during the extension reaction creating a fusion of two template strands.

115. Any of these errors can then be propagated during subsequent cycles of PCR, as a strand in which an error was introduced can be a template for a polymerase in a subsequent round. Due to the exponential nature of PCR, if a library preparation process contains multiple rounds of PCR and an error occurs early in the cycles of PCR, a large number of resulting strands may incorporate the error.

116. Likewise, SBS can be a source of error in the determined sequence. Each DNA fragment sequenced by SBS is amplified to form a small cluster with thousands of copies of the fragment on the surface of a flow cell (as in the Illumina SBS platforms) or even more copies on the surface of a bead (as in the 454 platform). Thus, at every cycle of sequencing when nucleotides are introduced, DNA

polymerase can potentially fail to incorporate a nucleotide into a subset of copies within a given cluster. For example, at the end of the third cycle of SBS on an Illumina sequencer, most of the DNA copies will have had their sequencing primers extended by three nucleotides, but a small fraction of them might be "behind" the others, having only incorporated two nucleotides. This phenomenon is known as "incomplete extension," which is a type of "dephasing." Dephasing can also result from "carry-forward," where some DNA copies spuriously incorporate more nucleotides than expected given the SBS cycle number. Dephasing errors accumulate within a cluster over time, and so at later cycles of sequencing, it can be difficult for the imaging system to correctly identify the nucleotide that was incorporated, resulting in sequencing errors.

117. From the beginning of NGS, scientists began looking for ways to avoid these sources of error. One solution was to simply avoid PCR entirely prior to sequencing. But this was not always practical, particularly where the concentration of input DNA was low. Furthermore, this does not eliminate errors occurring during the SBS process itself. Accordingly, scientists needed ways to associate PCR copies with an original template strand in a way that could correct for error in subsequent rounds of PCR or SBS. By 2012, one common way of doing this was to attach an identifying artificial sequence, sometimes referred to as a molecular tag, to the original fragments. Then, if mutations were observed in a particular sequencing

read, it would be possible to compare them to other reads with the same molecular tag sequence. For example, if 20 copies with the same tag (and therefore the same originating fragment) were sequenced and 18 of them had the same sequence but two of them appeared to contain a mutation, the mutation was likely the result of a PCR or SBS error.

118. The problem of differentiating between PCR copies was not unique to NGS—indeed it was a fundamental problem with the PCR technique, which generally did not preserve the ability to discriminate between an original template strand and copies of that strand. Accordingly, scientists developed the use of molecular tags even before NGS. In 2003, Hug and Schuler published a paper introducing the theoretical idea of using molecular tags "to count the number of individual molecules pertaining to the same molecular species." EX1059 (Hug & Schuler) at 0001-02. The next year, in 2004, scientists began putting theory into practice, describing the use of molecular tags, added prior to PCR amplification, to enhance sequencing accuracy. EX1021 (Miner) at 0001-04.

1) Hug & Schuler 2003

119. In 2003, Hug and Schuler recognized that it was then "not possible to label and count single mRNA molecules of a molecular species in a complex mixture." EX1059 (Hug & Schuler) at 0001. The authors described a "theoretical suggestion on how to count the number of individual molecules pertaining to the

same species in a complex molecular sample." EX1059 (Hug & Schuler) at 0002. The method described would have "random[ly] label[ed]" each molecule "with an oligonucleotide tag" that were "the same length but with different sequences of base pairs (tags)." EX1059 (Hug & Schuler) at 0002-03. The molecular tags would have been added by a PCR tail, as shown in the figure below. EX1059 (Hug & Schuler) at 0004-05. As the authors explained, "if the number of tags is sufficiently large, a unique tag will be attached" to each individual molecule from the sample. EX1059 (Hug & Schuler) at 0002.

120. Hug & Schuler's process of adding molecular tags is shown in the figure below, reproduced from EX1059 (Hug & Schuler) at 0005.

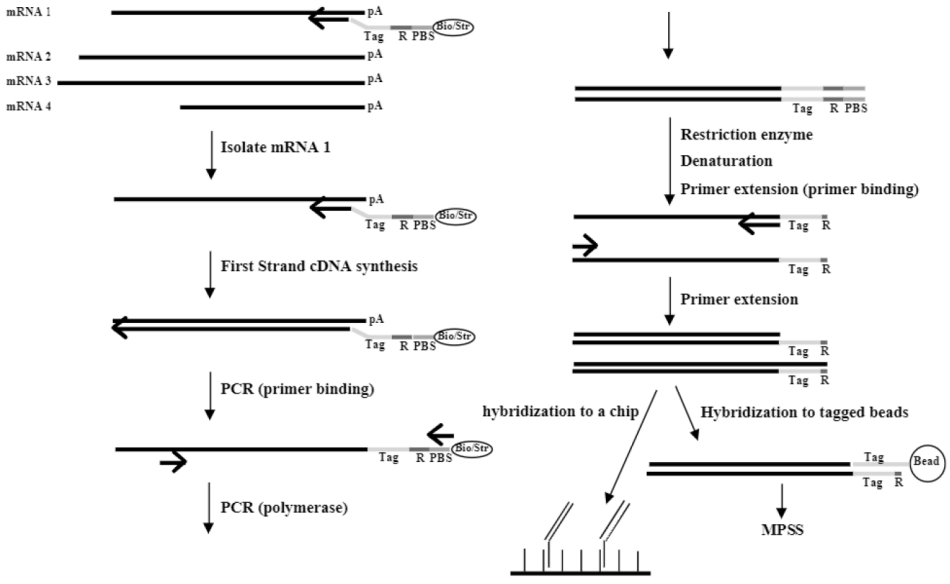


FIG. 1. Overview for the determination of the absolute number of mRNA molecules of a sequence of interest (IMPSMS) in a complex mRNA mixture by tagging.The molecules of mRNA1 are converted to new DMS. Only one tag sequence is shown. The steps shown are: hybridization of tag sequences to the total RNA or mRNA preparation, isolation of new DMS via paramagnetic streptavidine, first strand cDNA synthesis with reverse transcriptase, amplification by PCR, removal of PBS by restriction cutting, primer extension to with DNAse free Klenow polymerase to create double–stranded DNA leaving a single–stranded tag, (1) hybridization to tagged beads for further sequencing by MPSS or (2) direct hybridization to a DNA chip containing reverse complementary tag sequences. pA, polyA tail; Bio/Str, biotin/paramagnetic streptavidine; PBS, primer binding site; R, restriction endonuclease recognition site.

2) Miner 2004

121. In 2004, Miner et al. described a method of using molecular tags—which Miner called "molecular barcodes"—to "detect redundancy and contamination in hairpin-bisulfite PCR." EX1021 (Miner) at 0001–04. Miner incorporated a random 7-nucleotide sequence (the molecular tag) "in the hairpin linker sequence" which was ligated "to the ends of double-stranded genomic DNA fragment prior to bisulfite conversion and PCR amplification," as shown in the figure below from EX1021 (Miner) at 0001–02.

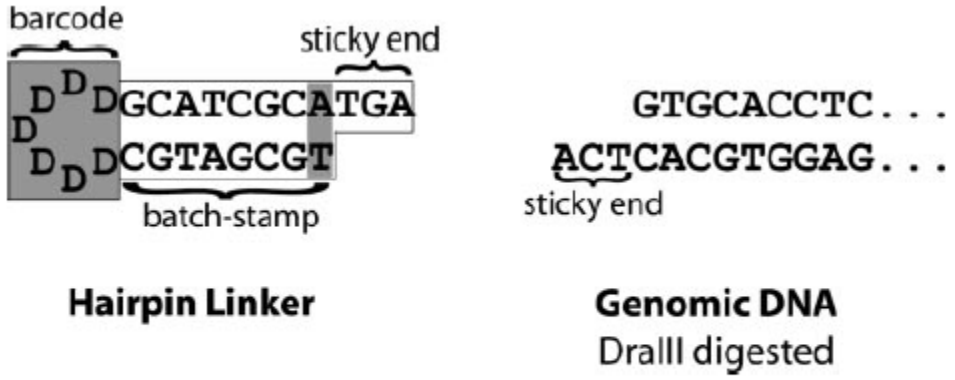


Figure 1. Schematic of barcoded and batch-stamped hairpin linker, designed for ligation to DraIII-cut genomic DNA of *FMR1*. The letter D represents a nucleotide randomly selected from A, G and T.

122. The molecular tag was "used to encode each ligated genomic fragment with information that distinguishes it from other sequences within a sample." EX1021 (Miner) at 0002. The molecular tag thus allowed Miner "to evaluate cloned sequences for redundancy and contamination," quantify original sample molecules, and correct "for skewing that can arise from PCR amplification or the cloning" of

the sample. EX1021 (Miner) at 0002–03. This method served "as a highly accurate method for positive identification of desired sequences." EX1021 (Miner) at 0002. In other words, Miner described a method of identifying and removing duplicate sequencing reads using molecular tags.

123. While Miner used molecular tags with hairpin-bisulfite PCR, Miner recognized that "the method can readily be adapted to single-strand PCR." EX1021 (Miner) at 0001.

3) McCloskey 2007

124. Several publications would later describe the development and use of molecular tags with other forms of PCR. In 2007, McCloskey et al. also incorporated a sequence of random nucleotides as a molecular tag, which McCloskey referred to as a "barcode," into an adapter to reduce PCR error. EX1022 (McCloskey) at 0001–07. McCloskey recognized that using PCR poses "risks of contamination, biased amplification, and product redundancy," and that "the reliability of PCR-derived assays would be significantly improved if methods were available to distinguish valid from data-corrupting sequences." EX1022 (McCloskey) at 0001–02.

125. McCloskey described a method using an adaptor for use with "single-stranded DNA or cDNA templates" that contained, among other elements, a 7-nucleotide long random molecular tag. EX1022 (McCloskey) at 0002. McCloskey incorporated the adaptor into a primer with a partially complementary region and a

noncomplementary tail. EX1022 (McCloskey) at 0002. When the primer was extended through the template strand, the resulting strand then had the noncomplementary region with the molecular tag at the 5′ end. McCloskey's molecular tag and process for appending it to a nucleic acid molecule is shown below in the figure reproduced from EX1022 (McCloskey) at 0003.

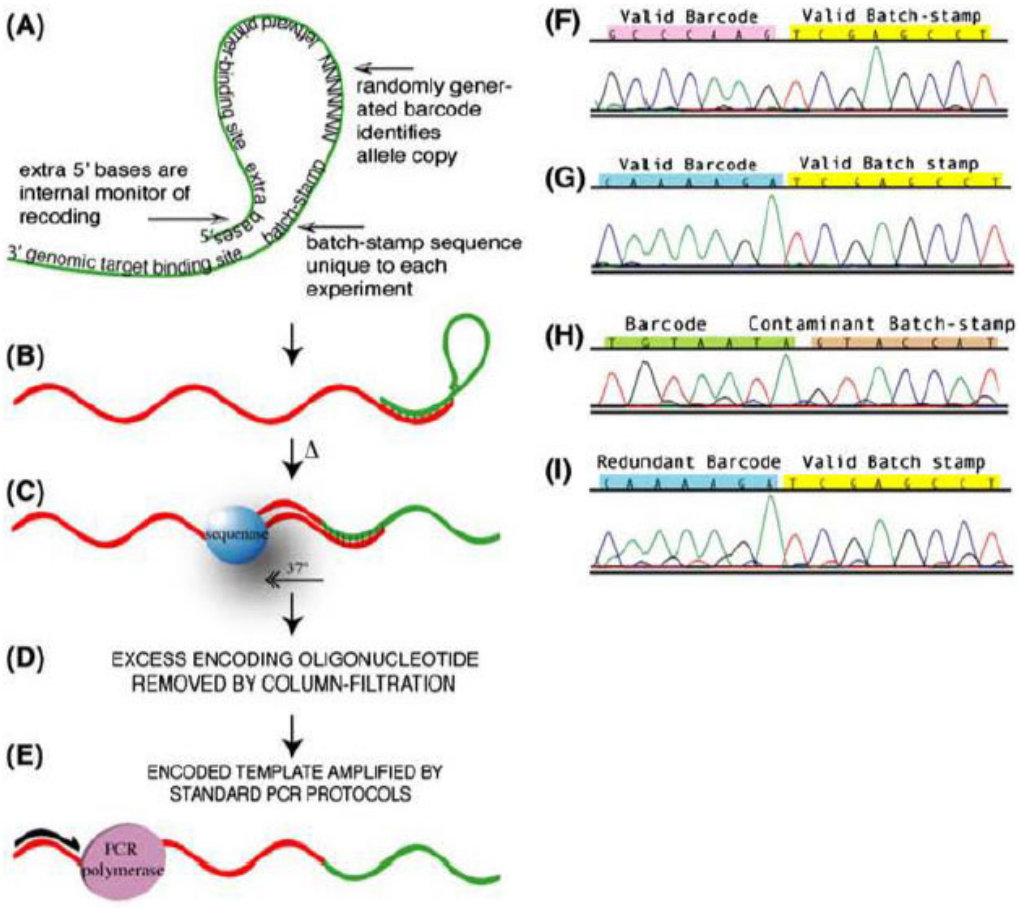


Fig. 1 Encoding oligonucleotide (**a**) has five distinct regions: a genomic target-binding site, a batch-stamp, a barcode, a primer-binding site for PCR, and extra bases that serve as internal sentinels of recoding as well as tethering the oligonucleotide to itself, thus minimizing nonspecific annealing. Encoding oligonucleotides are annealed to denatured genomic DNA for 15 min at 37°C (**b**) and extended at 37°C by Sequenase (**c**). The resulting encoded DNA strands are column-purified with Qiagen QIAquick columns (**d**) to remove unincorporated encoding oligonucleotides prior to PCR amplification (**e**). PCR products were sequenced and their encoded information was evaluated. Three classes of encoded sequences recovered from a single PCR amplification are illustrated here. Of 320 sequences analyzed, 299 were valid, each having the correct batch-stamp and a distinct barcode (**f** and **g**); 13 were contaminants, having batch-stamps that were used in previous reactions (**h**); and 8 were redundant, having barcodes and the batch-stamp identical to sequences previously analyzed from the same PCR (**i**)

126. McCloskey described the use of molecular tags to identify and remove PCR duplicates from sequencing data. With a molecular tag, "redundant sequences recovered from the same PCR reaction are readily identified because they are identical in both barcode and batch-stamp." EX1022 (McCloskey) at 0006. In other words, McCloskey taught a method of identifying and removing duplicate sequencing reads using molecular tags.

127. McCloskey's method had several practical applications. For example, "identifying redundant sequences and removing them from data sets is especially important when analyzing mosaic samples, which occur in both genetic and epigenetic diseases." EX1022 (McCloskey) at 0006. In all, McCloskey's incorporation of molecular tags prior to standard PCR amplification would have "markedly increase[d] the reliability of PCR-based sequence information" that had applications in several areas of study, including "forensic, clinical, single cell, and ancient DNA analyses." EX1022 (McCloskey) at 0007.

128. McCloskey also explained that molecular tags could be designed to be different numbers of nucleotides long. EX1022 (McCloskey) at 0004. Because only 1 of 4 bases is incorporated into a DNA strand at any one position, there are 4^n possibilities for uniquely labeling original sample molecules where "n" equals the number of nucleotides in the molecular tag. EX1022 (McCloskey) at 0004. For example, if one uses a molecular tag that is 7 nucleotides long, there are 4^7 (or

16,384) possible molecular tags. EX1022 (McCloskey) at 0004. Scientists would have understood McCloskey to be teaching that the molecular tag could be designed to be different lengths depending on the needs and objectives for any particular experiment.

129. In 2010, the number of publications describing the use of molecular tags increased considerably. That is because Illumina released its HiSeq 2000 sequencing instrument in 2010, which provided a substantial increase in sequencing throughput. The increased throughput made NGS significantly more practical and less expensive, and so scientists looked for ways to continue to improve upon the advancements in sequencing technology, including by using molecular tags to reduce PCR bias. Along with continuing to improve on the use of molecular tags incorporated into adapters, scientists in the field also started to explore using the sequence of the nucleic acid fragment from the sample as a unique identifier, or an endogenous "molecular tag."

4) Hiatt 2010

130. For example, in 2010, Hiatt et al. described the use of the sequence of the nucleic acid fragment as a method of identifying "groups of short reads sharing clonal origin" that could, post-sequencing and during data analysis, be reassembled into an accurate long sequencing read. EX1023 (Hiatt) at 0001–05.

131. In Hiatt's method, DNA was first sheared into approximately 500bp

fragments and those fragments were ligated with "tag-adjacent" adapters. Hiatt at 1. Those fragments were amplified by PCR. EX1023 (Hiatt) at 0001. The resulting amplified fragments were again sheared via sonication "and a single 'breakpoint-adjacent' adaptor was ligated to the sheared fragments." EX1023 (Hiatt) at 0001. Those fragments were then subjected to another round of PCR using tag-adjacent adapter primers and breakpoint-adjacent adapter primers. EX1023 (Hiatt) at 0001–02. The sequencing read adjacent to the tag-adjacent adapter "provides access to the genomic sequence that corresponds to the ends of the long fragments," and because that "sequence will be consistent across amplicons derived from the same long fragment, it can serve as a tag to identify molecules that are clonally derived." EX1023 (Hiatt) at 0002. Hiatt then used a "20-bp tag read and 76-bp breakpoint read" to reassemble the short reads into a single long read. EX1023 (Hiatt) at 0002. Hiatt's process is shown in the figure below, reproduced from EX1023 (Hiatt) at 0010.

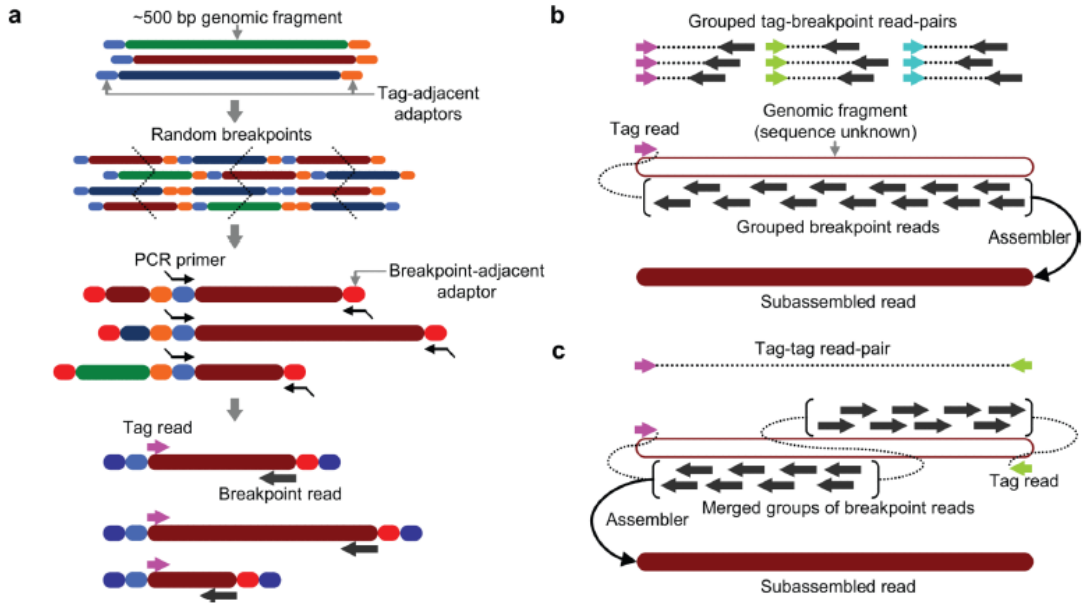


Figure 1.
Schematic of subassembly process. (a) Long DNA fragments are ligated to tag-adjacent adaptors, diluted and PCR-amplified. Dilution imposes a complexity bottleneck so that a limited number of long fragments are amplified. Concatemerized PCR products are then sheared by sonication and ligated to a breakpoint-adjacent adaptor. A second PCR amplification prepares amplicons for sequencing; one end of these amplicons corresponds to an end of a long fragment and the other end corresponds to a shearing breakpoint internal to that fragment. (b) Breakpoint reads are grouped in silico based on the sequence of the corresponding tag read. Breakpoint reads within a group, which derive from positions internal to the same parent long fragment, are subjected to local assembly to generate a subassembled read. (c) The metagenomic bottlenecked long-fragment library is subjected directly to paired-end Illumina sequencing to identify pairs of tag reads that were derived from opposite ends of the same original fragment. Two groups of breakpoint reads defined by distinct tag reads are merged and assembled together to generate one or more subassembled reads. In this study, this step was only applied to the metagenomic sample.

132. One potential problem of the method Hiatt identified was that the "tag sequences from different long DNA fragments can occasionally be identical by chance." EX1023 (Hiatt) at 0004. To solve for that problem, Hiatt suggested the use a "a simple modification." EX1023 (Hiatt) at 0004. The simple modification was "to use a tag-adjacent adaptor containing an embedded degenerate sequence (for example, a randomized 20-bp segment)" because that "would completely decouple the tag sequence from the sample composition." EX1023 (Hiatt) at 0004. In other

words, Hiatt described a method of identifying and removing duplicate sequencing reads using molecular tags.

5) König 2010

133. Also in 2010, König et al. developed a method of using molecular tags with cDNA to better understand protein-RNA interactions. EX1024 (König) at 0001–07. Specifically, König studied the role of heterogeneous nuclear ribonucleoprotein (hnRNP) C in splicing regulation. EX1024 (König) at 0001–02. UV-cross-linking and immunoprecipitation (CLIP) was already used as a method of determining protein-RNA interactions *in vivo*. EX1024 (König) at 0005. But CLIP suffered from a limitation "that most cDNAs truncate" before the cross-link nucleotide and so the sequence of that site was "lost during the amplification process." EX1024 (König) at 0005. König "developed individual-nucleotide resolution UV-cross-linking and immunoprecipitation (iCLIP)," where an adapter containing a molecular tag could be used to quantify cDNAs that truncate at the same nucleotide. EX1024 (König) at 0001–02. The process is shown in the figure below, reproduced from EX1024 (König) at 0011.

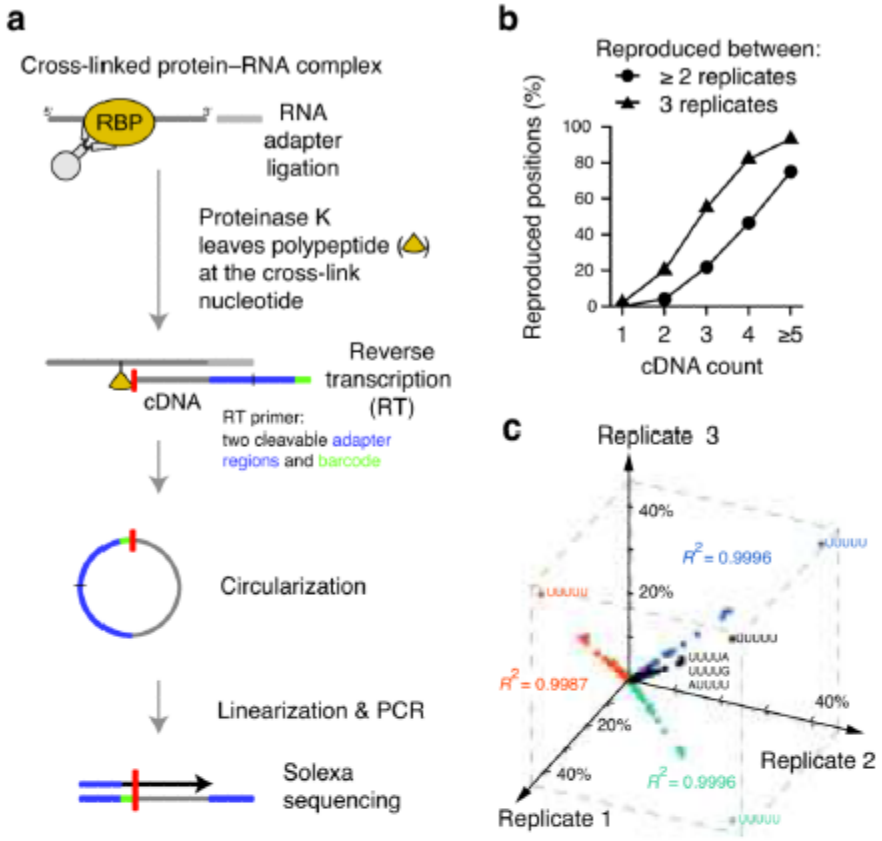


Figure 1.
iCLIP identifies hnRNP C cross-link nucleotides on RNAs. **(a)** Schematic representation of the iCLIP protocol. After UV irradiation, the covalently linked RNA is co-immunoprecipitated with the RNA-binding protein (RBP) and ligated to an RNA adapter at the 3′ end. Proteinase K digestion leaves a covalently bound polypeptide fragment on the RNA that causes premature truncation of reverse transcription (RT) at the cross-link site. The red bar indicates the last nucleotide added during reverse transcription. Resulting cDNA molecules are circularized, linearized, PCR-amplified and subjected to high-throughput sequencing. The first nucleotides of each sequence contain the barcode followed by the nucleotide where cDNAs truncated during reverse transcription. **(b)** Reproducibility of cross-link nucleotide positions. Percentage of cross-link nucleotides with a given cDNA count that were identified in at least two (circles) or all three experiments (triangles) are shown. The percentage of reproduced cross-link nucleotides increased with the incidence of hnRNP C cross-linking (cDNA count). **(c)** Reproducibility of sequence composition at cross-link nucleotides. Frequencies of pentanucleotides overlapping with cross-link nucleotides are shown for the three replicate experiments ($R^2 = 0.9996$, $R^2 = 0.9987$ and $R^2 = 0.9996$) with the sequence shown for the four most highly enriched pentanucleotides. 42% of cross-link nucleotides overlap with UUUUU in all three replicate experiments.

134. Specifically, König's method used a 3-nucleotide random molecular

tag—which König referred to as a "random barcode"—"to mark individual cDNA molecules." EX1024 (König) at 0005, 0007. By using molecular tags, König "eliminated PCR amplification artifacts by removing sequences that truncated at the same nucleotide in the genome and share the same random barcode." EX1024 (König) at 0002. The molecular tag thus allowed König "to discriminate between unique cDNA products and PCR duplicates." EX1024 (König) at 0002. In other words, König described a method of identifying and removing duplicate sequencing reads using molecular tags.

6) Kinde 2011

135. In 2011, recognizing that error rates for NGS were "generally too high to allow confident identification of rare variants," Kinde et al. described using molecular tags—which Kinde referred to as "unique identifiers" or "UIDs"—to "substantially increase the sensitivity" of NGS platforms. EX1025 (Kinde) at 0001. Identifying true genetic mutations had important applications in various fields of study, including oncology. EX1025 (Kinde) at 0001. As Kinde recognized, "counting genetic or epigenetic changes in tumors can inform fundamental issues in cancer biology." EX1025 (Kinde) at 0001. To accurately assess a genetic sequence it would have been necessary to reduce or eliminate errors introduced during sample preparation, including during PCR amplification. EX1025 (Kinde) at 0001.

136. Kinde developed "the Safe-Sequencing System" or "Safe-SeqS."

EX1025 (Kinde) at 0001. Broadly, Kinde's system proceeded in two steps. In the first step, "each DNA template molecule to be analyzed" was associated with a molecular tag. EX1025 (Kinde) at 0001. In the second step, the sample was amplified and analyzed. EX1025 (Kinde) at 0001. If a mutation was present in the original nucleic acid molecule from the sample, it should have been present in every copy produced during the amplification step. EX1025 (Kinde) at 0001. Under Kinde's method, if "at least 95%" of the same sequence contained the mutation it was identified as a true, natural mutation. EX1025 (Kinde) at 0001–02. In other words, Kinde taught a method of identifying and removing duplicate sequencing reads using molecular tags.

137. Like Hiatt in 2010, Kinde described methods of using the nucleic acid fragment as a unique identifier—an endogenous "molecular tag"—and a synthetic molecular tag added to an adapter during a PCR amplification step—an exogenous molecular tag. EX1025 (Kinde) at 0002–05. For the endogenous "molecular tag," Kinde explained that "randomly sheared genomic DNA inherently contains" unique sequences at each end of the sheared fragment. EX1025 (Kinde) at 0002. Upon reading the sequences, those fragments with identical sequences at the ends were considered to be from the same initial fragment from the sample. EX1025 (Kinde) at 0002. For exogenous molecular tags, Kinde introduced a 12-14-nucleotide molecular tag via two cycles of PCR, where the molecular tag was attached to the

tail of a primer. EX1025 (Kinde) at 0003. Kinde's exogenous tag process is shown in the figure below, reproduced from EX1025 (Kinde) at 0003.

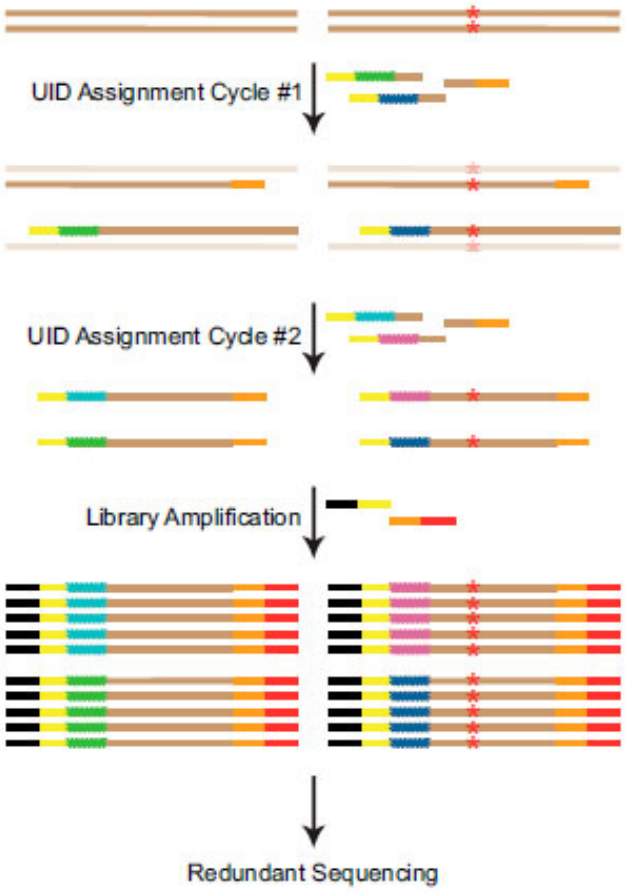


Fig. 3. Safe-SeqS with exogenous UIDs. DNA (sheared or unsheared) is amplified with a set of gene-specific primers. One of the primers has a random DNA sequence (e.g., a set of 14 Ns) that forms the unique identifier (UID) (variously colored bars), located 5′ to its gene-specific sequence, and both have sequences that permit universal amplification in the next step (yellow and orange bars). Two UID assignment cycles produce two fragments—each with a different UID—from each double-stranded template molecule, as shown. Subsequent PCR with universal primers, which also contain "grafting" sequences (black and red bars), produces UID families that are directly sequenced. Supermutants are defined as in the legend to Fig. 1.

138. Kinde reported that its method "substantially improve[d] the accuracy of massively parallel sequencing." EX1025 (Kinde) at 0004. The method could

therefore have been used to, among other things, "identify rare mutants in a population of DNA templates." EX1025 (Kinde) at 0004–05.

7) Casbon 2011

139. Also in 2011, Casbon et al. developed a method of using molecular tags to identify and remove PCR duplicates from sequencing reads to improve the accuracy of NGS. EX1026 (Casbon) at 0001–08. Like others, Casbon recognized that PCR amplification "changes the representation of the template molecules in the amplified product and introduces random error" and that it would be beneficial to remove PCR-introduced errors from sequencing data. EX1026 (Casbon) at 0001, 0004. Removal of PCR error would have enhanced NGS for several reasons, including by allowing the identification of "rare sequence variants" in "tumor samples." EX1026 (Casbon) at 0007.

140. Casbon described a method whereby a 13-nucletodide partially degenerate molecular tag—referred to as a "molecular counter," "degenerate base region," or "DBR"—was "ligated to all fragments during library preparation" prior to PCR amplification. EX1026 (Casbon) at 0001, 0003–06.

141. In one application reported by Casbon, the molecular tag is "used to determine whether a putative variant is associated with a single template molecule or, alternatively multiple template molecules and hence the probability that it derives from a polymerase error or true variant." EX1026 (Casbon) at 0001. Using

molecular tags "was highly effective at eliminating errors" and therefore improving the "detection of polymerase or sequencing errors," which reduces the chance of identifying a false variant. EX1026 (Casbon) at 0004-05. In another application, the molecular tag is used to count the number of original molecules. EX1026 (Casbon) at 0001. Because NGS methods require adapter ligation, Casbon also posited that molecular tags could "be incorporated into existing protocols at no extra cost in time and little extra cost in adapters." EX1026 (Casbon) at 0006. In other words, Casbon taught a method of identifying and removing duplicate sequencing reads using molecular tags.

142. Like McCloskey, Casbon also recognized that the molecular tag could be designed to be various nucleotides in length. EX1026 (Casbon) at 0001. Increasing the number of nucleotides in the molecular tag would increase the number of unique tags and thereby reduce, depending on the sample size, the possibility that two different fragments are tagged with the same molecular tag. EX1026 (Casbon) at 0005. But "for many applications it is only necessary to quantify low numbers of template molecules where miscalls can occur, in which case the number of DBRs can be set appropriately." EX1026 (Casbon) at 0005. In sum, the length of a molecular tag can be designed to account for experimental conditions.

8) Fu 2011

143. Also in 2011, Fu et al. reported a method of labeling nucleic acid

fragments with molecular tags. EX1027 (Fu) at 0001–06. Once the molecular tags were ligated onto the nucleic acid fragments, Fu explained that "the number of distinctly labeled targets" could be counted, and that doing so allowed the determination of "the number of molecules" in the original sample. EX1027 (Fu) at 0001. This principle is shown in Fu's figure below, reproduced from EX1027 (Fu) at 0001.

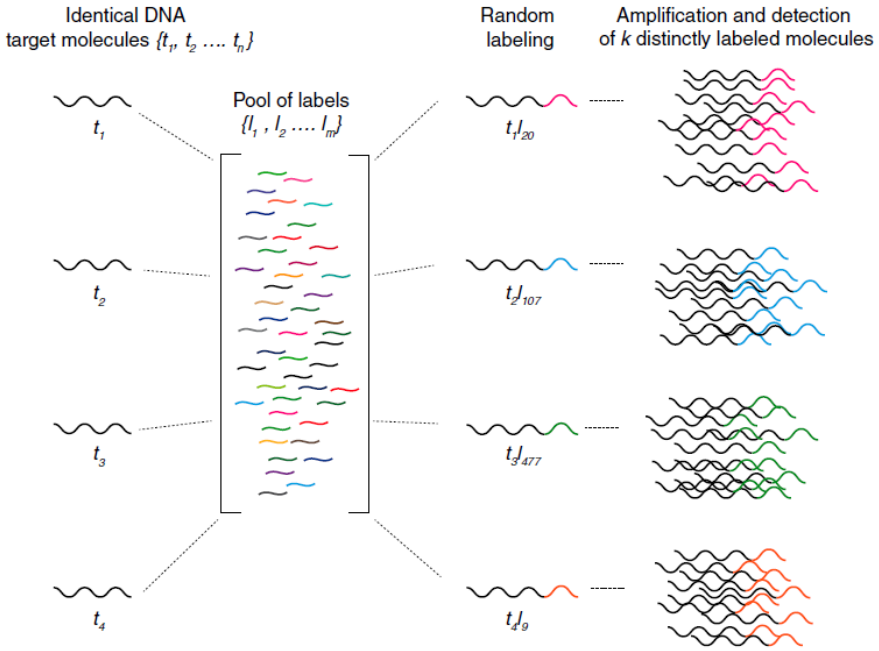


Fig. 1. A schematic representation of the labeling process. An example showing four identical target molecules in solution. Each DNA molecule randomly captures and joins with a label by choosing from a large, nondepleting reservoir of *m* labels. Each resulting labeled DNA molecule takes on a new identity and is amplified to detect the number of *k* distinct labels.

144. Specifically, Fu added 960 unique 14-nucleotide molecular tags to adapters that were added to a reaction mixture containing genomic DNA. EX1027 (Fu) at 0002, 0005. The adapters containing molecular tags were then randomly

ligated to the genomic DNA fragments. EX1027 (Fu) at 0002–03, 0005. Once the adapters were ligated, the sample was amplified by PCR and further processed. EX1027 (Fu) at 0003, 0005. This specific process is outlined in the figure below, reproduced from EX1027 (Fu) at 0002.

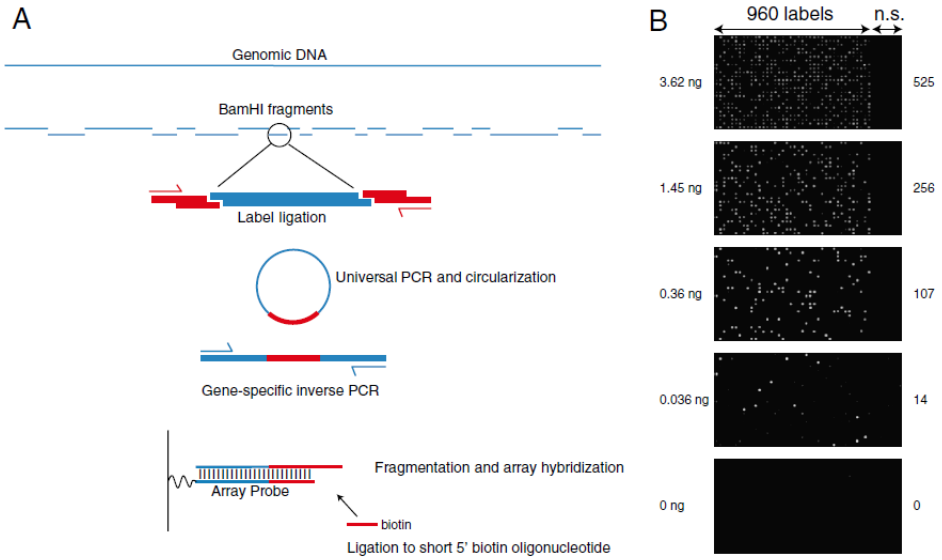


Fig. 3. (*A*) A schematic drawing of the method used to attach labels to fragments of DNA in the genome. Red bars represent a pool of synthetic deoxyoligonucleotide adaptors incorporating a collection of 960 labels used as counting sequences. A common primer sequence flanks each unique label adaptor, allowing universal amplification of all fragments with PCR. Circularization of amplified DNA molecules simplifies the selection and amplification of label-ligated DNA fragments through inverse PCR with gene-specific primers. The identity of labels that have been ligated to the genomic DNA fragment is determined using microarray hybridization, or DNA sequencing. (*B*) Microarray scan images of the 960 tiled probes for chromosome 4 corresponding to the labels used, as well as an additional 192 nonspecific (n.s.) probes serving as negative controls. The amount of genomic DNA used in each experiment is given on the left side of each image and the number of labels detected on microarrays is provided on the right side.

145. Using the molecular tags, Fu was able to take "a population of indistinguishable molecules," uniquely label them, and then count the number of original molecules in the sample. EX1027 (Fu) at 0004. The use of molecular tags to count original nucleic acid fragments in a sample had several applications, including "determining the absolute number of multiple target molecules in a specified container, such as high-sensitivity clinical assays, or for determining the

number of transcripts in single cells." EX1027 (Fu) at 0004. Counting original molecules using the molecular tag necessarily required the identification and removal of duplicate sequencing reads. In other words, Fu taught a method of identifying and removing duplicate sequencing reads using molecular tags.

9) Jabara 2011

146. Scientists also discovered that the applications of molecular tags extended to the study of RNA. In 2011, Jabara et al. recognized that while NGS had remarkable potential, it also had limitations. EX1028 (Jabara) at 0001, 0004–06. As discussed in other publications, the limitations principally came from the use of PCR prior to sequencing. EX1028 (Jabara) at 0001, 0004. Jabara discussed that PCR could cause several problems that inhibit the determination of natural nucleic acid sequences: (1) misincorporation of nucleotides by a polymerase; (2) artifactual recombination during amplification; (3) differential amplification, which "can skew allelic frequencies;" and (4) the creation of mass quantities of copied nucleic acids which can obscure the genetic variation in the original sample. EX1028 (Jabara) at 0001.

147. Jabara studied the use of molecular tags in viral RNA. EX1028 (Jabara) at 0001–06. To uniquely tag each molecule of viral RNA, which had to be converted to cDNA for sequencing, Jabara designed a cDNA synthesis primer that contained an 8-nucleotide molecular tag, which Jabara referred to as a "degenerate block of

nucleotides" or a "Primer ID." EX1028 (Jabara) at 0002. The molecular tag was introduced during the initial round of cDNA synthesis and so was "carried through all of the subsequent manipulations," including rounds of PCR. EX1028 (Jabara) at 0002. Jabara's process is shown in the figure below, reproduced from EX1028 (Jabara) at 0002.

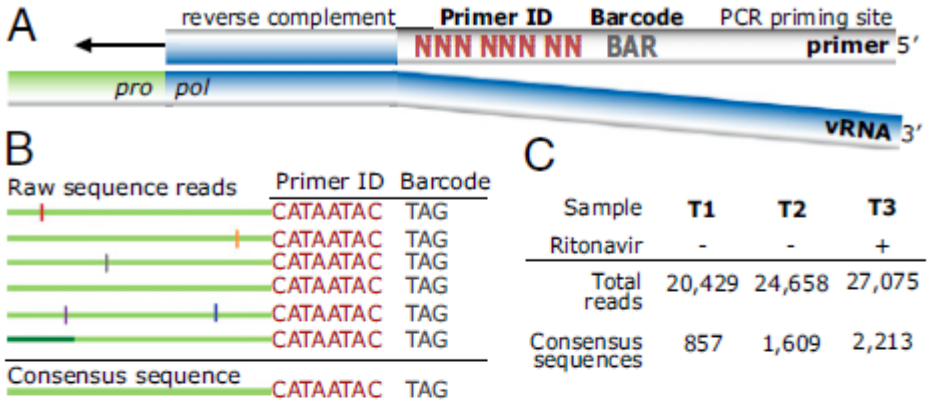


Fig. 1. Tagging viral RNA templates with a Primer ID before PCR amplification and sequencing allows for direct removal of artifactual errors and identifies resampling. (*A*) A primer was designed to bind downstream of the protease coding domain. In the 5′ tail of the primer, a degenerate string of eight nucleotides created a Primer ID, allowing for 65,536 unique combinations. An a priori selected three nucleotide barcode was designed for the sample ID. Finally, a heterologous string of nucleotides with low affinity to the HIV-1 genome was included in the far 5′ end for use as the priming site in the PCR amplification. (*B*) PCR biases and sequencing error are introduced during amplification and sequencing of viral templates. Repetitive identification of the barcode and Primer ID allow for tracking of each templating event from a single tagged cDNA. As errors are minor components within the Primer ID population, forming a consensus sequence directly removes them, and corrects for PCR resampling. (*C*) HIV-1 RNA templates isolated from plasma samples from two pre- and one postintermittent ritonavir drug therapy were tagged, amplified, and deep sequenced. Tagged sequences containing full-length protease were used to create a population of consensus sequences when at least three sequences contained an identical barcode and Primer ID.

148. Jabara reported that using molecular tags allowed for the identification and removal of duplicate sequencing reads: "Using this approach, we were able to

directly remove error, correct for PCR resampling, and capture the fluctuation of minor variants in the viral population within a host." EX1028 (Jabara) at 0002. Simply, Jabara's molecular tag allowed the determination of "a consensus sequence for each template, avoiding both PCR-related errors and sequencing errors." EX1028 (Jabara) at 0005. Jabara, moreover, recognized that molecular tags also allow for counting the original number of molecules in a sample: "In addition, the number of different Primer IDs reflects the number of templates that were actually sampled." EX1028 (Jabara) at 0005.

149. Jabara explained that the use of molecular tags could enhance our understanding of viral genomes, and particularly drug-resistant variants. EX1028 (Jabara) at 0001–02, 0005. But Jabara, citing Kinde, also recognized that the application of molecular tags extended to other nucleic acids, like DNA: "We believe a strategy that allows an initial tagging of individual templates before PCR and subsequent sequence analysis will be essential for understanding the true complexity and diversity of genetically dynamic populations." EX1028 (Jabara) at 0002, 0005–06.

10) Kivioja 2012 (online 2011)

150. Also in 2011, Kivioja et al. independently reported on the use of molecular tags. EX1012 (Kivioja) at 0001–03. Kivioja was made available online in 2011 in the journal *Nature Methods*. EX1012 (Kivioja) at 0001. I am familiar

with this journal from my work in the field of NGS. Like many papers in that journal, Kivioja was made available online prior to print. On the face of the article and on the *Nature Methods* website, it is indicated that the Kivioja paper was made available online on November 20, 2011. EX1012 (Kivioja) at 0001. In my experience, this is the date when the article would have been publicly available on the *Nature Methods* website.

151. As is customary for papers in *Nature Methods*, Kivioja was published along with associated supplemental materials available on the journal's website. The authors note this in the text of the article, when they state "[m]ethods and any associated references are available in the online version of the paper at" the *Nature Methods* website. EX1012 (Kivioja) at 0003. As is further customary, the published version of the article includes citations to the Supplementary Material and Supplementary Figures. Based on my past experience with this journal, this supplementary material would have been made available online on the same day as the rest of the article.

152. Kivioja teaches a method to "improve the accuracy of almost any next-generation sequencing method." EX1012 (Kivioja) at 0001. Kivioja disclosed in Figure 1, reproduced below from EX1012 (Kivioja) at 0001, multiple different ways of distinguishing original fragments from PCR copies, including labeling with molecular tags, fragmentation, and aliquoting.

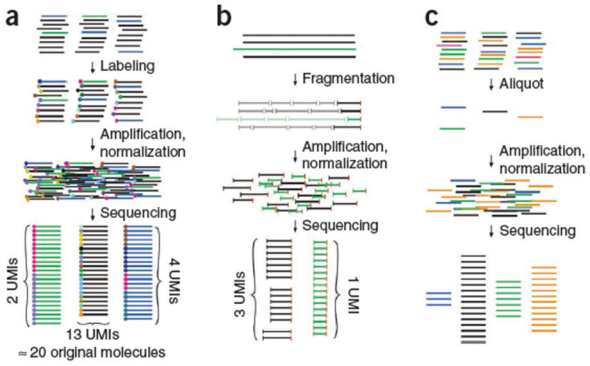


Figure 1 | UMIs can be generated by adding oligonucleotide labels, fragmenting, taking a small enough aliquot or a combination thereof. (**a**) Three different DNA species (green, blue and black lines) are labeled with a collection of random labels (colored filled circles). Two green molecules are originally 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.

153. As Kivioja explained "[a]ny 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 . . . ." EX1012 (Kivioja) at 0001. In other words, Kivioja explained that the fragmentation or aliquot methods allowed distinguishing fragments based on the fragment sequence alone, by ensuring that each fragment would be unique. But in other applications, Kivioja recognized that sequences could be distinguished based on the combination of a random label sequence and the fragment sequence, as shown in Figure 1(a).

154. As Kivioja explains, the UMIs (whether based on the fragment sequence alone or the fragment sequence and the label sequence together) can be used to "[d]etermin[e] the relative abundance of two different molecular species or the absolute number of molecules in a single sample," wherein the UMIs "act as a molecular memory of the number of molecules in the starting sample." EX1012 (Kivioja) at 0001.

155. Kivioja conducted two proof of concept experiments, one with DNA and one with mRNA. EX1012 (Kivioja) at 0002–03. For the experiment with DNA, much like Kinde, Kivioja recognized that the sequence of the nucleic acid fragment could be used as unique identifier. EX1012 (Kivioja) at 0002. Specifically, Kivioja reported that the "mapped genomic position of either end" of the fragmented DNA sequence "can be used" as a unique identifier. EX1012 (Kivioja) at 0002. With that, Kivioja was able to count the original number of DNA molecules in the sample. EX1012 (Kivioja) at 0002.

156. For the mRNA experiment, Kivioja designed an Illumina-compatible 5′ adaptor used during cDNA synthesis that contained a random 10-nucleotide molecular tag, as well as a sample index. EX1012 (Kivioja) at 0004, 0014. A partial adaptor was then added to the fragments during cDNA synthesis via reverse transcription using a template switch reaction. EX1012 (Kivioja) at 0004. A complete adapter was then formed by a subsequent PCR. EX1012 (Kivioja) at 0004. The adapter is suitable for use on an Illumina device, and comprises a primer binding site for the SBS reaction primer, a molecular tag, and a sample index. EX1012 (Kivioja) at 0001–02, 0014. As Kivioja explains, "The UMI method is compatible with sample indexing using separate DNA barcodes, allowing parallel analysis of samples." EX1012 (Kivioja) at 0014. Kivioja's adaptor and molecular tag are shown in the table below, reproduced from EX1012 (Kivioja) at 0014.

Supplementary Table 1. Sequences for oligomers used in mRNA-seq. Sequences in square brackets and parentheses represent an index sequence that was used to enable multiplexed sequencing, and a random ten base DNA sequence label, respectively. dU and rG represent deoxyuridine and guanine ribonucleotide, respectively.

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTVN

157. The POSA would also have understood that Kivioja necessarily added the P5 Illumina sequence, which is complementary to the sequence on the flow cell, in a subsequent step.

158. Kivioja used that molecular tag, along with the sequence of the nucleic acid fragment, to identify duplicate sequencing reads. EX1012 (Kivioja) at 0005. As Kivioja explained, "[t]he mapped reads with the same gene, position, and label were collected to one UMI and the number of such reads was recorded as the copy number of that UMI." EX1012 (Kivioja) at 0005. Once Kivioja identified and removed duplicate sequencing reads, Kivioja was able to count the number of original mRNA fragments in the sample. EX1012 (Kivioja) at 0003.

159. Kivioja's methods allowed the identification and removal of "read counting biases introduced by PCR," EX1012 (Kivioja) at 0003, and the POSA would have understood Kivioja's methods to also mitigate error from SBS. As Kivioja explained, the method therefore could "improve the accuracy of almost any

next-generation sequencing method." EX1012 (Kivioja) at 0001. In other words, Kivioja taught a generally-applicable method of identifying and removing duplicate sequencing reads for NGS sequencing using molecular tags.

11) Tollervey 2011

160. Also in 2011, Tollervey et al. used König's iCLIP protocol, described above, to "comprehensively identify the RNA sites interacting with TDP-43," a TAR DNA binding protein for which the "mislocalisation and aggregation is implicated in the pathogenesis" of certain neurological disorders. EX1029 (Tollervey) at 0001–02. As Tollervey (and König) explained, iCLIP "captures truncated cDNA and quantifies them using a random barcode"—the molecular tag—"contained within the primer used for reverse transcription." EX1029 (Tollervey) at 0002.

161. The molecular tag allowed for deduplication: "Duplicate sequences sharing the same random barcode were removed from analysis in order to avoid any apparent binding-site bias due to PCR artefacts. [sic]" EX1029 (Tollervey) at 0002. As did a combination of the molecular tag and part of the nucleic acid fragment: "After eliminating PCR amplification artefacts by removing sequences that truncated at the same nucleotide in the genome and shared the same random barcode, 3,7 million reads representing unique cDNA molecules were identified." EX1029 (Tollervey) at 0002. In other words, Tollervey taught a method of identifying and removing duplicate sequencing reads using molecular tags.

12) Shiroguchi 2012

162. Further development of molecular tags continued into 2012. In a paper of which I am an author, Shiroguchi et al. described the use of molecular tags with cDNA as a method of reducing PCR error reflected in sequencing data. EX1030 (Shiroguchi) at 0001–14. Shiroguchi was interested in transcriptome profiling, which requires accurately quantifying the abundance of RNA transcripts in a sample. EX1030 (Shiroguchi) at 0001. While PCR was not always necessary, for small amounts of sample material, it was. EX1030 (Shiroguchi) at 0001. As Shiroguchi explains, PCR introduces error and hinders that ability to accurately read the sequences of the original sample material. EX1030 (Shiroguchi) at 0001. To solve the PCR error problem, Shiroguchi designed 20-nucleotide molecular tags—which Shiroguchi referred to as "unique barcodes" or "barcodes"—that were introduced into double-stranded adapters that were then ligated to a cDNA library. EX1030 (Shiroguchi) at 0001–02, 0005. Shiroguchi's process and "unique barcodes" are shown in the figure below, reproduced from EX1030 (Shiroguchi) at 0002.

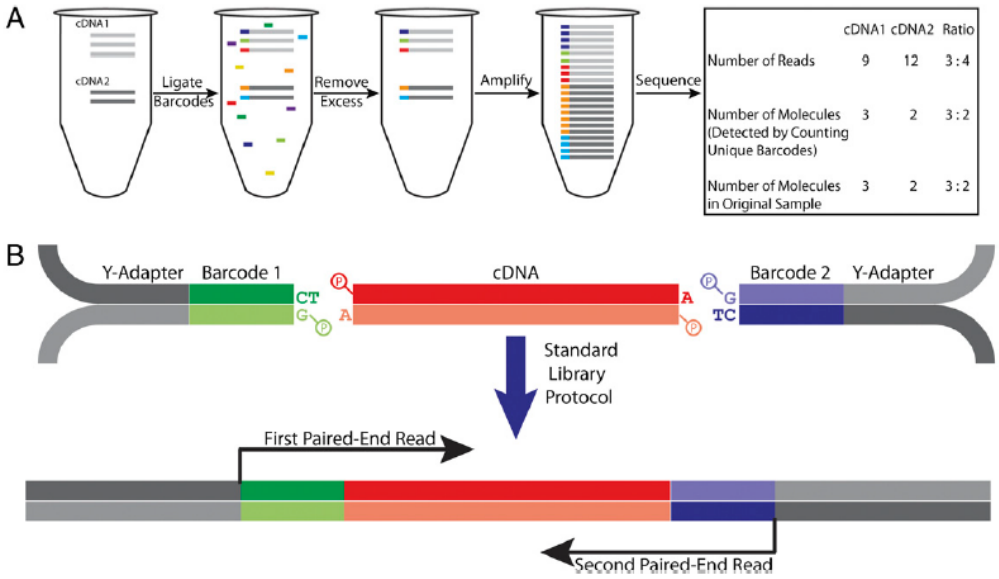


Fig. 1. Our scheme of digital RNA-Seq. (*A*) General principle of digital RNA-Seq. Assume the original sample contains two cDNA sequences, one with three copies and another with two copies. An overwhelming number of unique barcode sequences are added to the sample in excess, and five are randomly ligated to the cDNA molecules. Ideally, each cDNA molecule in the sample receives a unique barcode sequence. After removing the excess barcodes, the barcoded cDNA molecules are amplified by PCR. Because of intrinsic noise and sequence-dependent bias, the barcoded cDNA molecules are amplified unevenly. Consequently, after the amplicons are sequenced, it appears that there are three copies of cDNA1 for every four copies of cDNA2 based on the relative number of reads for each sequence. However, the ratio in the original sample was 3:2, which is accurately reflected in the relative number of unique barcodes associated with each cDNA sequence. (*B*) In our implementation of *A*, we found it advantageous to randomly ligate both ends of each phosphorylated cDNA fragment to a barcoded phosphorylated Illumina Y-shaped adapter. Note that the single T and A overhangs present on the barcodes and cDNA, respectively, are to enhance ligation efficiency. After this step, the sample is amplified by PCR and prepared for sequencing using the standard Illumina library protocol. For each amplicon, both barcode sequences and both strands of the cDNA sequence are read using paired-end deep sequencing.

163. As shown in the figure above, Shiroguchi incorporated molecular tags into adapters that were ligated to each end of the cDNA fragment. EX1030 (Shiroguchi) at 0002. Shiroguchi used the double-stranded molecular tags containing pre-defined sequences, which were read using paired-end sequencing, to identify PCR duplicates. EX1030 (Shiroguchi) at 0002, 0005. The use of double-stranded molecular tags enhanced PCR error correction because each fragment could be identified by not one but two unique tags. EX1030 (Shiroguchi) at 0002, 0004–05. Once Shiroguchi identified PCR duplicates using the molecular tags, the sequences were "aligned to the first 28 bases (26 bases for the second sequencing run) of the targeted sequence." EX1030 (Shiroguchi) at 0005. Only sequences that

had expected molecular tags and aligned with the expected sequence of the nucleic acid fragment (with the allowance of 2-nucleotide mismatches) were kept. EX1030 (Shiroguchi) at 0005. Shiroguchi therefore described a method of identifying and removing duplicate sequencing reads using molecular tags.

13) Schmitt 2012

164. Also in 2012, Schmitt et al. reported the use of double-stranded 12-nucleotide molecular tags. Schmitt's method first required incorporating single-stranded random 12-nucleotide tags into "standard Illumina sequencing adapters." EX1031 (Schmitt) at 0001–09. Using a DNA polymerase, Schmitt then generated a complementary strand to the random 12 nucleotides. EX1031 (Schmitt) at 0002, 0006–08. This resulted in a double-stranded molecular tag, which Schmitt referred to as a "Duplex Tag." EX1031 (Schmitt) at 0002. The adapters were then ligated to DNA fragments and PCR amplified. EX1031 (Schmitt) at 0002, 0006–08. Schmitt depicted the molecular tag and process for appending it to the nucleic acid fragment in the figure below, reproduced from EX1031 (Schmitt) at 0002.

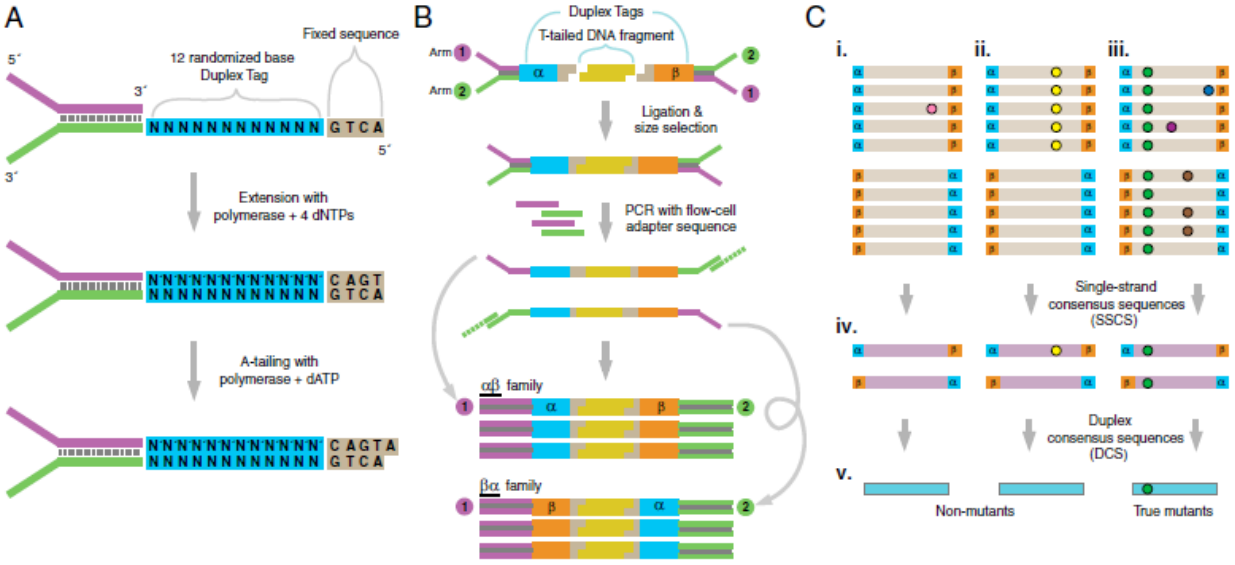


Fig. 1. Overview of Duplex Sequencing. (*A*) Adapter synthesis. A double-stranded, randomized Duplex Tag sequence is appended to a sequencing adapter by copying a degenerate sequence in one strand of the adapter with DNA polymerase. Complete adapter A-tailing is ensured by extended incubation with polymerase and dATP. (*B*) Duplex Sequencing workflow. Sheared, T-tailed double-stranded DNA is ligated to A-tailed adapters. Because every adapter contains a Duplex Tag on each end, every DNA fragment becomes labeled with two distinct tag sequences (arbitrarily designated α and β in the single fragment shown). PCR amplification with primers containing Illumina flow-cell–compatible tails is carried out to generate families of PCR duplicates. Two types of PCR products are produced from each DNA fragment. Those derived from one strand will have the α tag sequence adjacent to flow cell sequence 1 and the β tag sequence adjacent to flow cell sequence 2. PCR products originating from the complementary strand are labeled reciprocally. (*C*) Error correction. (*i–iii*) Sequence reads sharing a unique set of tags are grouped into paired families with members having strand identifiers in either the αβ or βα orientation. Each family pair reflects the amplification of one double-stranded DNA fragment. (*i*) Mutations (colored spots) present in only one or a few family members represent sequencing mistakes or PCR-introduced errors occurring late in amplification. (*ii*) Mutations occurring in many or all members of one family in a pair arise from PCR errors during the first round of amplification such as might occur when copying across sites of mutagenic DNA damage. (*iii*) True mutations (green) present on both strands of a DNA fragment appear in all members of a family pair. Whereas artifactual mutations may co-occur in a family pair with a true mutation, all except those arising during the first round of PCR amplification can be independently identified and discounted when producing (*iv*) an error-corrected single-strand consensus sequence (SSCS). The sequences obtained from each of the two strands of an individual DNA duplex can then be compared to obtain (*v*) the duplex consensus sequence (DCS), which eliminates remaining errors that occurred during the first round of PCR.

165. Schmitt then used the molecular tags to identify duplicate sequencing reads: "After sequencing, members of each PCR family are identified and grouped by virtue of sharing an identical tag sequence." EX1031 (Schmitt) at 0002. Schmitt then collapsed the duplicate sequencing reads into a single sequencing read—that is, deduplicated them: "The sequences of uniquely tagged PCR duplicates are then compared to create a PCR consensus sequence. . . . This step filters out random errors introduced during sequencing or PCR to yield a set of sequences, each of which derives from an individual molecule of single-stranded DNA." EX1031 (Schmitt) at 0002.

166. Schmitt further reduced the presence of PCR error in final sequencing reads by using the molecular tag's complementary strand. EX1031 (Schmitt) at 0002. After identifying the single-stranded consensus read, Schmitt paired the single-stranded consensus reads with their corresponding complementary molecular tag sequence. EX1031 (Schmitt) 0002. In practice, this meant each original nucleic acid fragment was identified using a 24-nucleotide molecular tag (12 in the single-stranded read and 12 in its complement). EX1031 (Schmitt) at 0002. In the final read sequence, Schmitt retained a "sequence base at a given position" "only if the read data from each of the two strands matches perfectly." EX1031 (Schmitt) at 0002, 0006.

167. By comparing the sequence of a single strand and its complement, "each of the two strands in a duplex facilitates differentiation of sequencing errors from true mutations." EX1031 (Schmitt) at 0002. If a change in the sequence is present in both the single strand and its complement, then it can be assumed that the change is a true mutation. EX1031 (Schmitt) at 0002. But if a change is present in only one strand, it can be assumed that it is PCR error. EX1031 (Schmitt) at 0002. In other words, Schmitt taught a method of identifying and removing duplicate sequencing reads using molecular tags.

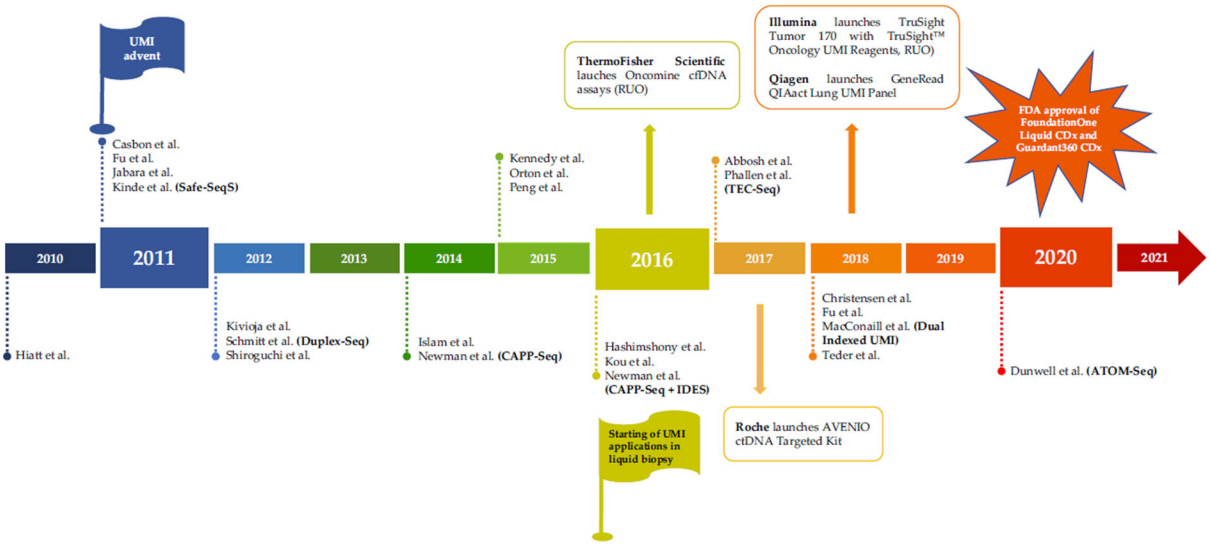
168. Like other molecular tag methods described, Schmitt's method solved the problem of identifying "errors introduced during sample preparation and

sequencing." EX1031 (Schmitt) at 0001. The ability to use molecular tags to detect minor variants had practical clinical applications, including the possible "early detection of cancer." EX1031 (Schmitt) at 0001.

169. Schmitt also recognized that molecular tags could be incorporated into various adapters and methods that were applicable across a wide range of NGS platforms. EX1031 (Schmitt) at 0005. The method would have been "generalized to essentially any sequencing platform" as the molecular tag could be, among other things, "incorporated into existing adapters." EX1031 (Schmitt) at 0005.

14) Later Publications

170. As a recent review recognized, the use of molecular tags has drastically improved the capability of NGS methods to identify rare genetic alterations. EX1032 (De Luca) at 0003–05. The use of molecular tags was developed much earlier than NGS methods were. EX1032 (De Luca) at 0003. So scientists were primed to apply molecular tags to NGS. Indeed, from 2010 to 2012 and later, several groups developed and reported on the use of molecular tags in NGS to identify and remove PCR duplicates and thereby reduce PCR error. EX1032 (De Luca) at 0004–05. A timeline of the development of molecular tags is depicted below and reproduced from EX1032 (De Luca) at 0005.



3. Single-Read and Paired-End Sequencing

171. Single-read sequencing is a method that sequences adaptor-nucleic acid fragments from one end of the fragment. All NGS platforms, including 454, Ion Torrent, and Illumina, allow for single-read sequencing.

172. Paired-end sequencing enabled the sequencing of adaptor-nucleic acid fragments from both ends. EX1009 (Bronner) at 0024. On the Illumina platform, paired-end sequencing takes advantage of bridge amplification. Each fragment is read first in one direction. Then, a strand complementary to the original fragment is generated and read from the other direction on the complementary strand. These "paired-end" strands are then associated with each other during data processing.

173. Paired-end sequencing has several benefits. NGS platforms in 2013 generated relatively short sequence reads. EX1014 (Masoudi-Nejad) at 0053. The fidelity and accuracy of DNA polymerases used in the SBS reactions also decreased

as the read length increased, and the quality of the read correspondingly decreased. So it was often beneficial to sequence fragments from both ends because the sequencing data could be compared, and thus in essence double-checked, or linked together to determine the sequence of a long fragment.

174. In addition, if the fragments were ligated to adaptors with sample indexes or molecular tags, paired end sequencing permitted experimental designs in which the index and/or tag were read on one end of the fragment, and the fragment sequence was read from the other side of the fragment. The POSA would have understood there were benefits to this arrangement—for example, the number of SBS reactions or cycles for the first read could be small, resulting in just enough bases to read the index and/or tag, and then the number of cycles for the second read from the other side of the fragment could be as many as the device permitted, resulting in a longer second read. This would have the effect of maximizing the length of the fragment sequenced while not losing any sequencing data to the ligated adaptors.

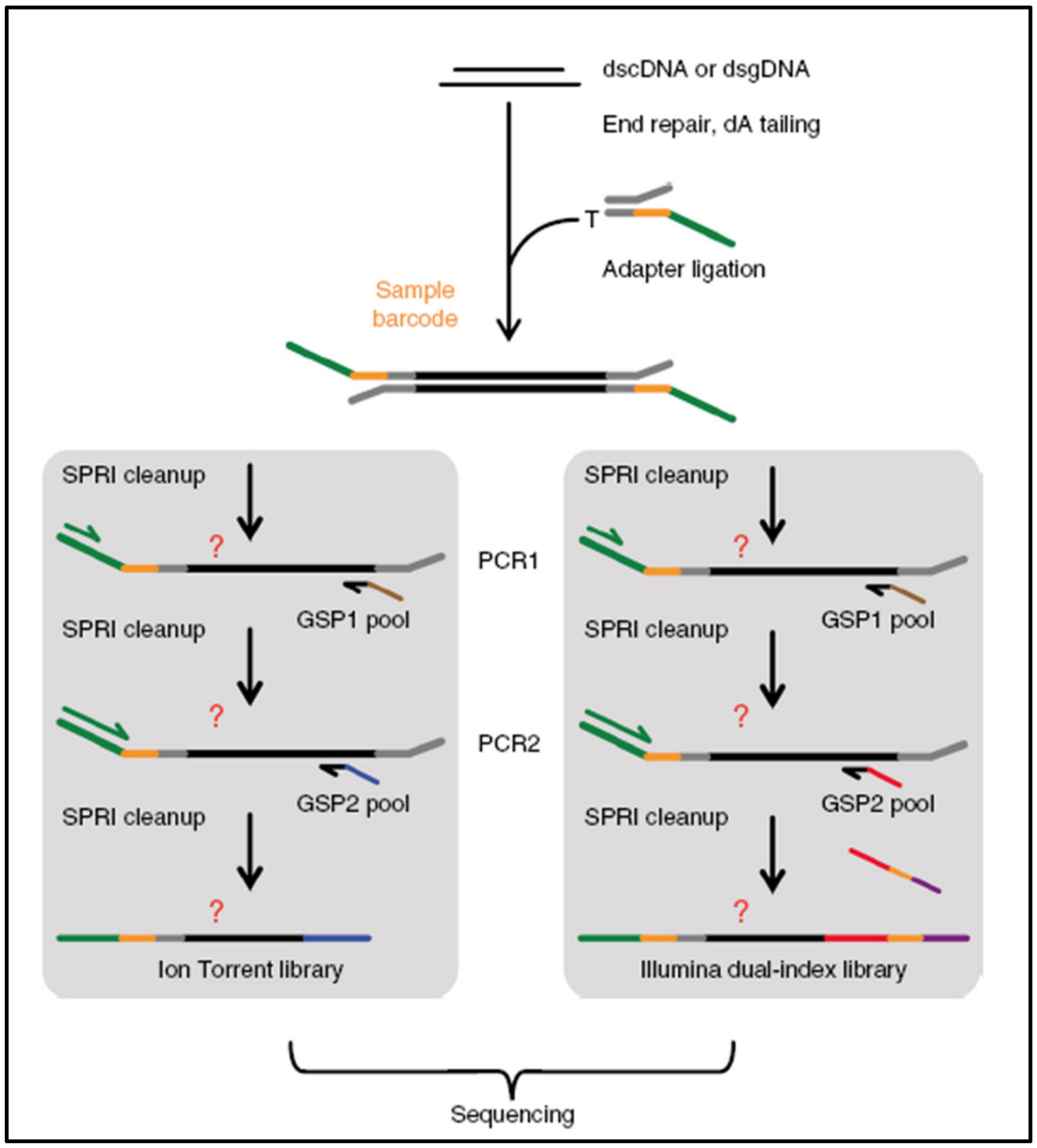
175. In the early 2010s, I routinely sequenced nucleic acid fragments using single-read and paired-end sequencing with the Illumina platform. In my experience, scientists in the field routinely optimized the arrangement of sequences in the adaptors and the use of paired-end sequencing depending on the particular experimental design, including the size of fragments expected, the length of the

various adaptor sequences, and the read length desired for the particular target sequences.

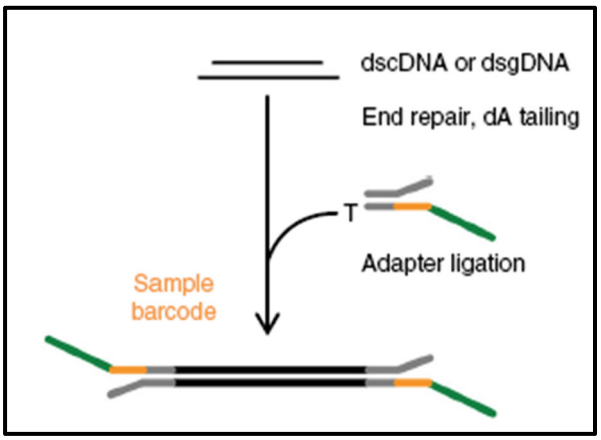
L. The MGH/Archer "Anchored Multiplex PCR" Library Preparation Method

176. In 2014, pathologists at Massachusetts General Hospital ("MGH") published an article in Nature Medicine detailing a "rapid target enrichment method" ("the MGH Method") they had developed for Next Generation Sequencing that built on early anchored-PCR approaches. EX1015 (Zheng). The MGH Method, which Zheng et al. referred to as "anchored multiplex PCR for targeted next-generation sequencing" was designed to detect unknown genetic mutations located adjacent to known genetic sequences. EX1015 (Zheng) at 0001. These unknown mutations included "gene rearrangements (without prior knowledge of the fusion partners), single nucleotide variants, insertions, deletions and copy number changes." EX1015 (Zheng) at 0001.

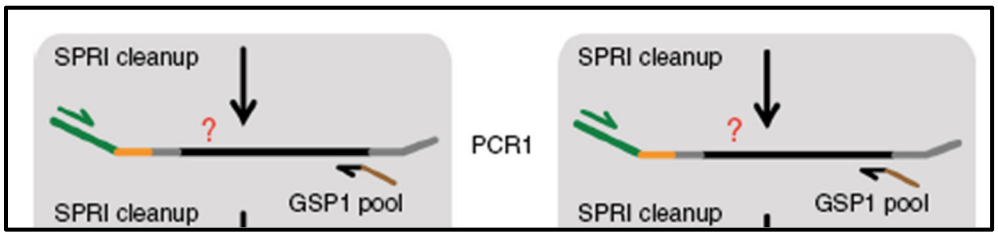
177. As shown in the below figure, reproduced from EX1015 (Zheng) at 0002, the MGH Method involves two rounds of PCR using two pools of "nested" gene-specific primers, "GSP1 pool" and "GSP2 pool."



178. As explained in Zheng, the MGH Method begins by ligating a "half functional adapter" to a double-stranded DNA fragment, comprised of either double-stranded genomic DNA ("dsgDNA") or double-stranded cDNA transcribed from RNA ("dscDNA"), as shown below in the figure reproduced from EX1015 (Zheng) at 0002.

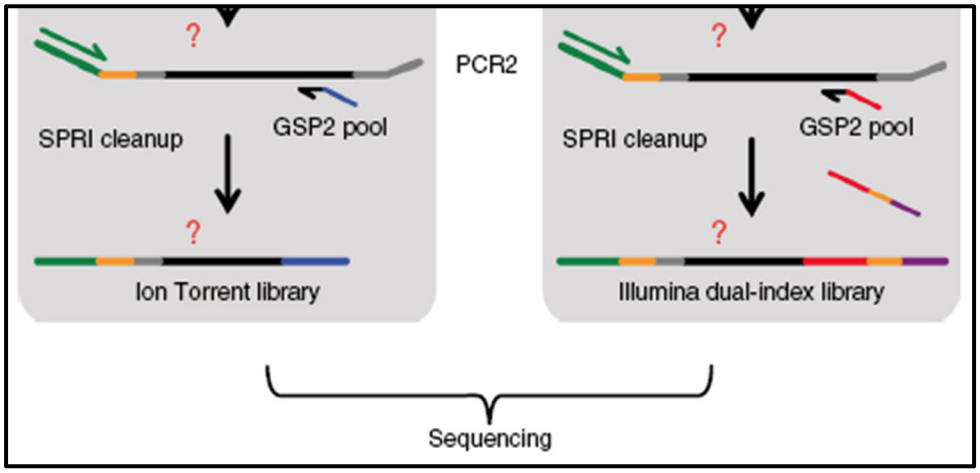


179. As shown in the figure below, reproduced from EX1015 (Zheng) at 0002, the adaptor-ligated fragment then undergoes a first PCR reaction ("PCR1") involving 10-14 cycles of PCR using an adaptor-specific primer (green, below), and a first set gene-specific primers ("GSP1 primers," black and brown, below). Each GSP1 primer has a sequence at its 3′ end that is complementary to a known genetic sequence adjacent to an unknown fusion partner or target sequence (designated with a question mark below), and a non-complementary sequence at its 5′ end. EX1015 (Zheng) at 0002.



180. The sample then under undergoes a second PCR reaction ("PCR2"), involving 10 cycles of PCR using an adaptor-specific primer (green, below), and a second set of gene-specific primers ("GSP2 primers"), which are depicted below as

black and blue (left), and black and red (right), for the Ion Torrent and Illumina platforms, respectively. Each GSP2 primer has a sequence at its 3′ end that is complementary to a known genetic sequence adjacent to an unknown fusion partner or target sequence that is "nested" (that is, located closer to the unknown sequence) relative to the GSP1 target-specific sequence. Each GSP2 primer also has a non-complementary sequence at its 5′ end containing a universal adaptor sequence specific for Ion Torrent or Illumina. In the Illumina version of the MGH method, a third primer is added during PCR2 containing a sequence complementary to the sequence of the tail of the GSP2 that includes a second indexing sequence at its own tail. The resulting library products are then sequenced on an Ion Torrent or Illumina device, as shown in the figure below reproduced from EX1015 (Zheng) at 0002.



181. I understand that the MGH method was commercialized as a library preparation kit for NGS by a company called ArcherDX and research-use-only ("RUO") versions of the kits embodying the MGH method are now manufactured

and sold by Integrated DNA Technologies, Inc. ("IDT").

182. I have created the below simplified diagram to summarize the nested PCR technique employed in the MGH method, showing the two cycles of PCR, the first involving a first pool of GSPs, and the second involving a second pool of GSPs that are "nested" relative to the first set of GSPs.



V. THE PERSON OF ORDINARY SKILL IN THE ART

183. As explained above in Section III.B of my declaration, I have been instructed by counsel that the POSA is a hypothetical individual with ordinary skill in the relevant field who is deemed to have knowledge of the relevant prior art as of the priority date of the patent.

184. As the breadth of the prior art shows, as of 2012 and 2013, NGS was still a nascent technology but was growing rapidly in popularity in multiple different academic disciplines. The POSA with respect to methods of library preparation for NGS would have had academic training in any of a variety of fields that used NGS, including chemistry, molecular biology, biochemistry, or medicine. The POSA would have had a Ph.D. or M.D. in one of these fields and several years of work experience. The POSA further could have had less academic experience and more practical experience.

185. In addition to their academic background and years of work experience, the POSA with respect to all of the patents would have had experience in implementing and designing library preparation methods for the leading NGS platforms at the time, including 454 and Illumina. This would have included specific experience with, and understanding of, the sequences contained in adaptors for NGS sequencing in 2012 and 2013, and the chemistry used to attach those adaptors to DNA fragments prior to sequencing.

186. At the relevant times in 2012 and 2013, I possessed these qualifications. I received my Ph.D. from Harvard in 2009 and completed my postdoctoral research there from 2009 through 2011, before starting my laboratory at Columbia in 2012. During my Ph.D., postdoctoral work, and my early years at Columbia, I worked extensively with NGS, including developing novel NGS technology and NGS library preparation techniques for both 454 and Illumina platforms. I published on this work in several publications during the relevant time period, including Sims, P.A., Greenleaf, W.J., Duan, H., Xie, X.S., *Fluorogenic DNA Sequencing in PDMS Microreactors*, NATURE METHODS, 8, 575–580 (2011); Shiroguchi, K., Jia, T.Z., Sims, P.A., Xie X.S., *Digital RNA Sequencing Minimizes Sequence Dependent Bias and Amplification Noise with Single Molecule Barcodes*, PROC. NATL. ACAD. SCI. USA, 109, 1347–1352 (2012).

VI. THE ’399, ’357, AND ’241 PATENTS

187. In this section, I provide my opinions for why the claims of the ’399, ’357, and ’241 patents were anticipated and/or would have been obvious to the POSA at the priority date. I have reviewed the file history for each of these patents and none of the references I rely on herein were considered by the Examiner during prosecution. These patents share the same specification but have different claims (though the claims are similar, as I discuss in this section).

188. The ’399 patent has two independent claims, claim 1 and claim 18. I

understand that IDT is challenging only claim 1 and claims that depend from that claim in this proceeding. Claim 1 is reproduced below:

> **1**. 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.

189. In addition to claim 1, I understand that IDT is challenging claims 2–3, 5–11, and 14–17 of the ’399 patent. For the reasons I explain below, claims 1–3, 5–11, and 14–17 would have been anticipated and/or obvious to a person skilled in the art by November 13, 2013, the ’399 patent’s priority date.

190. Claim 1 is the ’357 patent’s only independent claim, which I understand IDT is challenging along with certain claims that depend from claim 1. Claim 1 is

reproduced below:

> The invention claimed is:
> **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.

191. In addition to claim 1, I understand that IDT is challenging claims 2–4, 6–7, and 9–10. For the reasons I explain below, claims 1–4, 6–7, and 9–10 would have been anticipated and/or obvious to a person skilled in the art by November 13, 2013, the '357 patent's priority date.

192. Claim 1 is the '241 patent's only independent claim, which I understand IDT is challenging along with certain claims that depend from claim 1. Claim 1 is reproduced below:

> The invention claimed is:
> **1**. A method for detecting duplicate sequencing reads, the method comprising:
> obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor, wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment;
> sequencing the amplicons to generate sequence reads that include identifier and target sequences; and
> identifying sequence reads with identical identifier and target sequences as duplicates.

193. In addition to claim 1, I understand that IDT is challenging claims 2–16. For the reasons I explain below, claims 1–16 were anticipated and/or obvious by November 13, 2013, the ’241 patent’s priority date.

A. The Priority Date For The ’399, ’357, and ’241 Patents

194. I have been asked to assume for purposes of my opinions and analysis in this declaration that the priority date for the analysis of the ’399, ’357, and ’241 patents is November 13, 2013.

B. Claim Terms Defined In The ’399, ’357, and ’241 Patents.

195. The ’399, ’357, and ’241 patents share a specification.

196. I understand from counsel that the terms in a patent generally have the ordinary meaning that they would have to the POSA. But I further understand that where the patentee has provide an express definition for a particular term, that definition controls.

197. Where relevant terms are defined in the patents, I have applied those definitions in reaching my opinions in this declaration. Where particular terms are defined that are relevant only to a handful of particular claims, I discuss those definitions in those relevant sections.

198. For certain terms that appear frequently across multiple claims of the three patents, I have set forth those definitions in this section for convenience. While I have cited to the ’399 patent in this section, the same definitions appear in all three

patents. I apply these definitions throughout this declaration.

199. The patents define "indexing site" as "a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing." EX1003 ('399 patent) at 4:34–38. The POSA would understand that this definition is synonymous with what I discuss herein as a "sample index."

200. The patents define "identifier site" as "a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads." EX1003 ('399 patent) at 4:43–45. The POSA would understand that this definition is synonymous with what I discuss herein as a "molecular tag."

201. The patents define "target sequence primer binding site" as a "nucleotide sequence for binding a primer for a target sequence." EX1003 ('399 patent) at 4:62–63. The POSA would understand that this definition is consistent with a forward read primer binding site for an SBS reaction on an NGS device, e.g., an Illumina Genome Analyzer II or Illumina MiSeq, such as the binding site for a Nextera read primer. The patents provide one exemplary sequence for the "target sequence primer binding site" (SEQ ID No. 3): TCTTTCCCTACACGACGCTCTTCCGATCT. EX1003.0020 ("Illumina forward sequencing priming site"); EX1004.0023 ("forward priming site compliment" [sic]); EX1003.0023 ("forward priming site compliment" [sic]). The POSA would

recognize this sequence to be a conventional Illumina sequence known in the prior art. EX1056 (Illumina Letter) at 0005 ("D501-D508 adapters").

202. The patents define "indexing primer binding site" as a "nucleotide sequence for binding a primer for an indexing site." EX1003 ('399 patent) at 4:33–34. In the context of an Illumina device, this could refer to multiple different sites depending on the context. For example, in the context of single read sequencing, this definition could refer to the forward read primer binding site. But at the priority date, the POSA also would have recognized, as I discuss above, that in some Illumina devices there was a separate index read that used a primer that bound to a different site from the forward read primer binding site. Both of these sites would fall within this definition. The patents provide one exemplary sequence for the "Indexing read priming site" (SEQ ID No. 2): AGAGCACACGTCTGAACTCCAGTCAC. EX1003.0020 ("Indexing read priming site"); EX1004.0023 ("priming site"); EX1005.0023 ("priming site"). The POSA would recognize this sequence to be a conventional Illumina sequence known in the prior art. EX1056 (Illumina Letter) at 0005 ("D701-D712 adapters"). The patents otherwise direct the POSA to use sequences compatible with an Illumina device. EX1003 at 12:36-39 ("The methods of the present invention are useful for sequencing by the method commercialized by Illumina"), 16:44-52 ("an Illumina indexing read priming site such as AGAGCACACGTCTGAACTCCAGTCAC (SEQ ID NO:2)"); EX1004 at 12:51-

52, 16:60-63; EX1005 at 12:38-40, 16:46-48.

203. The patents define a "universal" primer as one "that can be used for amplifying a number of different sequences." EX1003 ('399 patent) at 5:57–58.

C. The Challenged Claims Of The '399, '357, And '241 Patents Would Have Been Obvious Over Iafrate And Kivioja.

204. In this section of my declaration I provide the reasoning for my opinion that claims 1–3, 5–11, 14–17 of the '399 patent; claims 1–4, 6–7, and 9–10 of the '357 patent; and claims 1–16 of the '241 patent would have been obvious and/or anticipated by Iafrate and Kivioja by November 13, 2013.

1. Iafrate Is Prior Art.

205. Iafrate was filed on March 11, 2013 and published on November 14, 2013 as U.S. Patent Application No. 2013/0303461. EX1011 (Iafrate). I have been informed by counsel that Iafrate is prior art as March 11, 2013, the date that it was filed at the Patent Office. Thus, I understand that Iafrate is prior art to the '399, '357, and '241 patents, which all claim priority to a provisional patent application that was not filed until November 13, 2013.

2. Kivioja Is Prior Art.

206. Kivioja is a *Nature Methods* article that published online in 2011 and in print in 2012. EX1012 (Kivioja) at 0001. I understand that Kivioja is prior art as of its 2011 online publication date. Thus, I understand that Kivioja is prior art to the '399, '357, and '241 patents, which all claim priority to a provisional patent

application that was not filed until November 13, 2013.

207. I am familiar with *Nature Methods* from my training and experience, and I was familiar with this journal in 2011 at the time Kivioja was published online. As is customary in the field, and in *Nature Methods* in particular, Kivioja refers to supplementary materials and methods available online, and states "methods and any associated references are available in the online version of the paper" at the *Nature Methods* website. EX1012 (Kivioja) at 0003. In my experience with scientific journals generally, and *Nature Methods* specifically, supplemental materials are made available online at the same time an article publishes online. Consistent with that practice, throughout the article, the authors of Kivioja cite directly to the supplementary materials and methods that were made available online, and reference to the supplemental material is necessary to fully understand the article. I therefore refer to Kivioja and its supplementary materials and online methods as a single reference ("Kivioja") throughout this declaration.

3. The POSA Would Have Been Motivated To Use Iafrate's Method and Adapter for NGS Library Preparation.

208. Iafrate discloses "methods of determining oligonucleotide sequences" by NGS that include "enriching target sequences prior to sequencing." EX1011 (Iafrate) at 0001. Iafrate explains that "[t]arget 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." EX1011 (Iafrate) at 0012 (¶5). To that end, Iafrate discloses "methods of enriching target sequences prior to a sequencing step" and in particular, methods wherein "the sequence of one end of the target sequence to be enriched is not known prior to the sequencing step." EX1011 (Iafrate) at 0014 (¶34). In other words, Iafrate teaches a method of NGS library preparation whereby a target nucleic acid sequence of interest is selectively enriched prior to sequencing. EX1011 (Iafrate) at 0012 (¶¶5–7), 0018–23 (¶¶74–112). The POSA would have understood that selective enrichment is the amplification of nucleic acid sequence of interest, generally by PCR, so as to increase the concentration of the fragments containing the sequence of interest relative to other DNA in the sample.

209. As Iafrate explains, the method is particularly useful for applications involving "a disease associated with a genetic alteration" such as cancer. EX1011 (Iafrate) at 0013 (¶15). This is because, as Iafrate explains, cancer is frequently associated with gene translocations and rearrangements, where a known oncogene appears at an unexpected place in the genome. EX1011 (Iafrate) at 0013–14 (¶¶15–25). Iafrate provides a proof of concept of the method, demonstrating that it can be used to detect mutations in cancer oncogenes like ROS1, ALK, and RET. EX1011 (Iafrate) at Figs. 2–7, 0014 (¶¶27–31), 0029 (¶233), 0030–31 (¶¶241–247). As the

inventors explain, "[u]sing this three gene target panel assay with multi-sample barcoding, ROS1, ALK and RET rearrangement positive samples have been detected using one Ion Torrent sequencing run." EX1011 (Iafrate) at 0031 (¶246).

210. In addition to the Iafrate method's usefulness in determining genetic alterations associated with cancer, Iafrate also explains that the method of target enrichment provides "high coverage depth" that "enables wider dynamic range for allele counting (in gene expression and copy number assessment)." EX1011 (Iafrate) at 0029 (¶227). The POSA would have understood that having the capability to, for example, assess copy number variants was also important to understanding genetic influences on cancer development. EX1074 (Lonigro) at 0001–11. Because of its applications to cancer research, the POSA would have been particularly interested in Iafrate's method and in using it for NGS library preparation. Because of the genetic origins of cancer, as of the priority date, cancer researchers and pathologists were particularly interested in harnessing the newly available NGS technology to detect and identify oncogenic mutations. In addition, the Iafrate methods are not limited to oncology applications because the POSA would recognize that they could be used for target enrichment for other areas as well, such as to sequence known variable sequences in the genome adjacent to fixed sequences, such as in immune cells. For the reasons Iafrate explains, including lower cost and the ability to multiplex more samples at once compared to shotgun

sequencing, the POSA would be interested in implementing Iafrate's method for NGS applications.

211. To accomplish selective enrichment, Iafrate appends an adaptor to one end of the nucleic acid fragment. EX1011 (Iafrate) at 0011, 0012–13. One such adaptor that Iafrate disclosed was a customized Illumina adaptor that incorporates all but one of the features in the adaptor recited in claim 1 of the '399 patent. EX1011 (Iafrate) at 0011; EX1003 at 25:39–46. Iafrate disclosed its adaptor in Figure 9, which I have reproduced below with my own annotations.[2] The Iafrate adaptor includes (i) an indexing primer binding site (referred to as a "barcode sequencing primer binding site") (in **blue** below), (ii) an indexing site (referred to as a "barcode") (in **orange**), and (iii) a target sequence primer binding site (referred to as a "sequencing primer binding site") (in **green**), as shown in the annotated version of EX1011 (Iafrate) Figure 9 I have included below. EX1011 (Iafrate) at 0011, 0012–13.

[2] Throughout this section I have added figures that are included the '399 patent and which I annotated for illustration purposes. In some places I have also added figures, which are not included in the '399 patent, for illustration purposes.

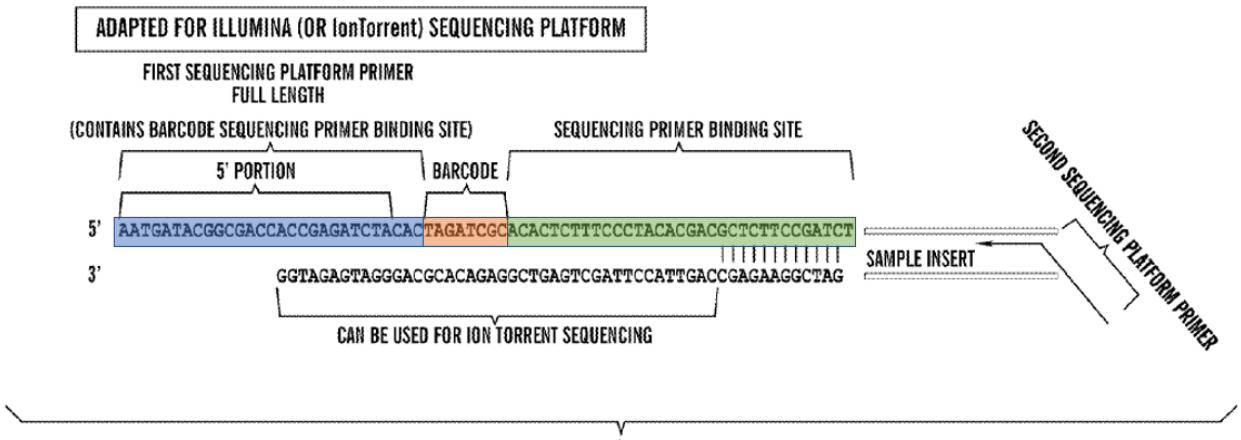


FIG. 9

212. Iafrate's method can be used with different types of NGS platforms, but Iafrate teaches that Illumina is the preferred platform, particularly when sequencing requires "higher accuracy." EX1011 (Iafrate) at 0029. In order to be compatible with Illumina devices that were on the market by 2013, Iafrate incorporated standard Illumina sequences in the adaptor disclosed in Figure 9. Indeed, Kircher et al., in Figure 1B, disclosed identical sequences at the positions of the indexing primer binding site and sequence primer binding site shown in Iafrate Figure 9. EX1055 (Kircher) at 0002. Kircher labeled the indexing/barcode primer binding site as the "P5 grafting" sequence and the sequence primer binding site as the "forward read primer." EX1055 (Kircher) at 0002. This would have been consistent with the POSA's understanding at the priority date that the Iafrate adaptor used a standard structure for an Illumina adaptor with the sequence complementary to the sequence on the chip (the indexing primer binding site / P5 grafting sequence) followed by a sample index, followed by the forward read primer binding site for reading the

fragment sequence in the sequencing reaction. As I explained in Section IV.K.2.a of my declaration, the POSA would have understood that, depending on the Illumina device used, the P5 grafting sequence also functioned as the indexing primer binding site. The POSA would have further recognized that these Illumina sequences could be rearranged, subject to certain constraints imposed by the device—for example, the P5 grafting sequence must appear at the 5′ end so as to permit the fragments to anneal to the grafted oligos on the surface of the flow cell.

213. Iafrate Figure 9 also labels one site as a sample index. The POSA would have recognized this sequence as a standard Illumina sample index, as evidenced by contemporaneous documentation. EX1056 (Illumina Letter) at 0003; EX1057 (Nextera Prep Guide) at 0044. Those sample index sequences were referred to as "Index 2" and/or the "i5" index because that sample index was added to the P5 side of the Illumina adapter, as shown in Kircher, Figure 1B. EX1055 (Kircher) at 0002.

214. The sample index primer binding site disclosed in Iafrate's Figure 9, which the POSA would have understood to also be the Illumina P5 graft sequence, allows the adaptor, and the nucleic acid fragment the adaptor is appended to, to anneal to the oligonucleotides bound to the Illumina solid support. It also allows the binding of an Illumina primer for the SBS reaction for the sample index read. The sequence primer binding site disclosed in Iafrate's Figure 9 allows the binding of an

Illumina primer for the SBS reaction for the sequencing read.

215. The POSA would have understood that Iafrate's preference for using adaptors compatible with Illumina devices is consistent with the general trend in the art, by 2013, towards a preference for sequencing on Illumina devices. EX1009 (Bronner) at 0001.

216. Iafrate further discloses that its method is suitable for sample pooling and sequencing, also known as multiplex sequencing. Iafrate teaches that the "barcode" (sample index) in Figure 9, EX1011 (Iafrate) at 0011, "allow[s] for multiplex sample pooling" because the nucleic acid fragments in a sample are "ligated with unique barcode adapters," EX1011 (Iafrate) at 0029.

217. In Iafrate, and as was generally known in the art at the time, the use of a sample index allows nucleic acid fragments derived from one particular sample (e.g., Sample A) to be sequenced with nucleic acid fragments derived from a different sample (e.g., Sample B). Once the nucleic acid fragments from one or more samples are sequenced, the sample index allows a particular sequencing read to be associated with its sample of origin (e.g., Sample A or B). In other words, the sample index, which is unique to each sample, allows sequencing reads to be demultiplexed.

218. Thus, as recited in the '399 patent, the POSA would have known that Iafrate's sample index would have been "unique amongst a subset of the plurality of nucleic acid fragments and [would have been] an index for multiple

polynucleotides." EX1003 ('399 patent) at 25:57–59. Indeed, Iafrate's sample index is equivalent to the "indexing site" recited in the '399 patent, EX1003 ('399 patent) at 25:43, which is defined as "a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing," EX1003 ('399 patent) at 4:34–38. The POSA also would have understood that the purpose of adding a sample index to an adaptor is to allow for multiplex sequencing.

219. Iafrate also teaches each of the library preparation and sequencing steps recited in claim 1 of the '399 patent. In claim 1(a), the '399 patent requires "ligating an adaptor to a 5′ end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples." EX1003 ('399 patent) at 25:39–41. Iafrate ligated adaptors to the 5′ end of nucleic acid fragments from one or more samples. EX1011 (Iafrate) at 0023, 0051. In Figure 9, which I have reproduced and annotated below, Iafrate shows the adaptor appended to the 5′ end of the nucleic acid fragment (referred to as the "sample insert"). EX1011 (Iafrate) at 0011.

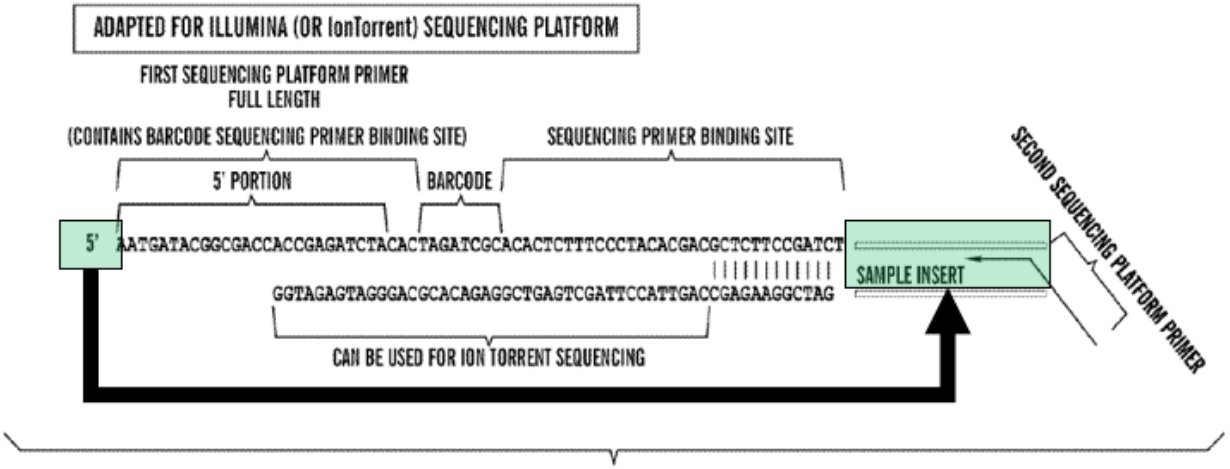


FIG. 9

220. The POSA would have understood that the adaptor in Figure 9 would have been ligated to the 5′ end of the nucleic acid fragment, especially in light of the fact that Iafrate taught ligating adaptors, including:

- "In step (a) of the method described herein, the universal oligonucleotide tail-adaptor can be ligated to the target nucleic acid." EX1011 (Iafrate) at 0017.
- "In some embodiments, a universal oligonucleotide tail-adaptor with a unique barcode portion can be added to each sample and ligated to the nucleic acids therein; the samples can then be pooled after step (a)." EX1011 (Iafrate) at 0023.
- "[A] randomly sheared, double-stranded cDNA or gDNA template can be end-repaired, adenylated, and then ligated on one or both ends with the universal Y adapter to create a universal sequence for initiating PCR and subsequent sequencing." EX1011 (Iafrate) at 0029. The POSA would have understood that ligating an adaptor "on one or both ends" of the nucleic acid fragment includes ligation on the 5′ end of the nucleic acid fragment.

221. Claim 1(b) requires "amplifying the adapter-nucleic acid fragment ligated products." EX1003 ('399 patent) at 25:47–48. Iafrate teaches, and claims,

that the adaptor-ligated nucleic acid fragments are amplified prior to sequencing. EX1011 (Iafrate) at 0051, 0029. The POSA would have also understood that amplification prior to sequencing would have been a routine step during NGS sample preparation.

222. Claim 1(c) requires "generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products." EX1003 ('399 patent) at 25:49–51. Iafrate taught that the amplified adaptor-ligated nucleic acid fragments would have been sequenced on an NGS device, including on an Illumina device: "Importantly, this method is adaptable for any other next generation sequencing platform, including but not limited to Illumina, SOLiD, and 454." EX1011 (Iafrate) at 0029, 0051. Indeed, Iafrate specifically taught that the adapter disclosed in Figure 9 was compatible for NGS on an Illumina device, as shown in my annotated version of the figure below. EX1011 (Iafrate) at 0011, 0029.

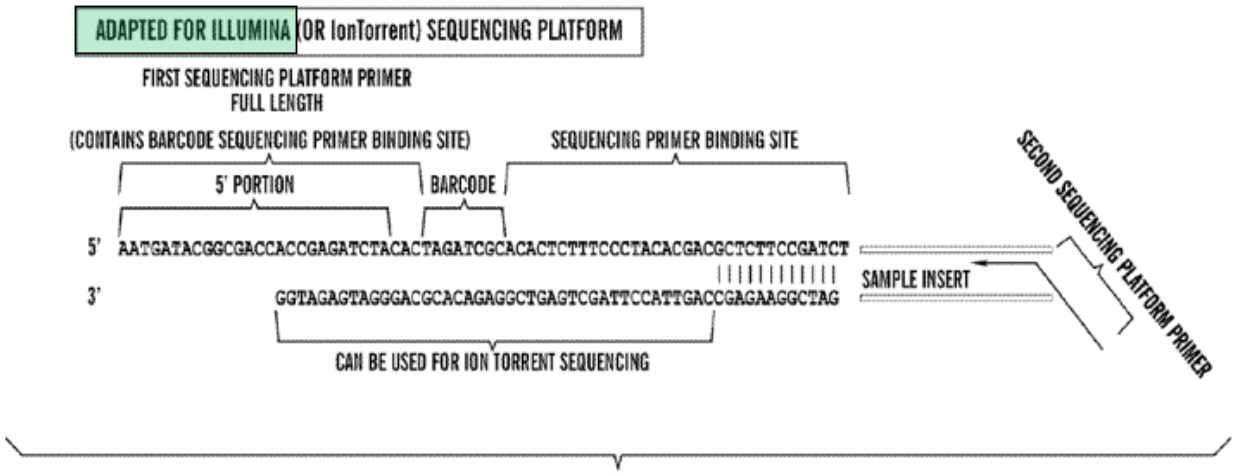


FIG. 9

223. Finally, the POSA would have understood that sequencing on an Illumina device would have "generat[ed] a population of sequencing reads," as the '399 patent recites. EX1003 ('399 patent) at 25:49.

224. Thus, Iafrate teaches nearly every element recited in claim 1 of the '399 patent. Specifically, Iafrate teaches ligating adaptors—comprising an indexing primer binding site, an indexing site, and a target sequence primer binding site—to the 5′ end of a plurality of nucleic acid fragments from one or more samples, after which the adapter-ligated nucleic acid products are amplified and sequenced. The only limitations Iafrate does not disclose are the use of a molecular tag in the adaptor and the identification of duplicate sequencing reads after an NGS run. But as I discuss below, it would have been obvious to the POSA to add these features, and the POSA would have been motivated to do so to improve Iafrate's sequencing method and would have had a reasonable expectation of success in doing so.

4. The POSA Would Have Been Motivated to Modify Iafrate's Method To Incorporate Molecular Tags And Use UMIs As Taught By Kivioja.

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.

226. By 2013, scientists knew that to understand genetic influences on cancer development and progression, and to identify potential therapeutic targets, they needed to identify unknown nucleic acid sequences adjacent to genes that are known to be biomarkers for particular tumors. EX1011 (Iafrate) at 0023–25. Indeed, at that time, scientists knew the development of cancer was associated with mutations like translocations and gene fusions where a known gene appears at an unexpected place in the genome. EX1011 (Iafrate) at 0023–25. Iafrate's method facilitates the identification of those unknown sequences, and thus the identification of potentially oncogenic gene fusions or translocations. EX1011 (Iafrate) at 0012, 0029–30.

227. Iafrate's method also makes it more likely to detect a particular sequence, such as low-frequency mutations, which Iafrate explains is "a critical feature for evaluating somatic mutations in cancer." EX1011 (Iafrate) at 0029. The capability of Iafrate's method to detect low-frequency mutations comes from the selective pre-sequencing PCR amplification of the nucleic fragments of interest. The

result of selective PCR amplification is that the amplified fragments containing the sequences of interest vastly outnumber other fragments that were present in the original sample. So by amplifying the target sequences prior to NGS, it would have been more likely that the POSA would have been able to identify low-frequency mutations that would not necessarily have been picked up and sequenced in shotgun sequencing.

228. The same process of selective pre-sequencing PCR amplification also explains why the Iafrate method can be used to enhance "allele counting (in gene expression and copy number assessment)." EX1011 (Iafrate) at 0029 (¶227).

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.

230. These amplification cycles confound the ability to use the sequencing results for "allele counting" or "copy number assessment" as Iafrate proposes. This is because it is very difficult after PCR to determine whether two identical observed

sequences are PCR copies of one another, or two different fragments from the original sample that happen to have the same sequence. This means that it is difficult to accurately count original fragments, particularly after the number of cycles of PCR described in Iafrate.

231. In addition, the POSA would have understood that each round of PCR can cause errors in the sequence of the nucleic acid fragment. Errors could result, for example, from misincorporation of nucleotides by a polymerase, the sequencing-by-synthesis process, non-specific primer binding, recombination, or incomplete extension of primers. Due to the exponential nature of PCR, errors that are introduced in an early cycle of PCR could be propagated. The earlier an error is introduced, the more it can skew the final sequencing data. In some circumstances, nucleic acid fragments with errors introduced during PCR could outnumber fragments that contain the naturally occurring sequence. Such an outcome would make it difficult to determine the naturally occurring sequence of that nucleic acid fragment or count the number of molecules in a sample. Indeed, the POSA would have understood that PCR-introduced errors would have made it more difficult to determine, for example, whether a mutation identified in sequencing data was present in the original sample from a tumor cell or was a product of PCR error.

232. Library preparation was only one of two sources of error. The POSA would have recognized that the Illumina device itself used multiple rounds of PCR

and also used a polymerase for the SBS reaction. Thus, the POSA also would have been concerned about, and would have sought to reduce, error introduced by the SBS process itself.

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.

234. In addition, for oncology applications like Iafrate's, the POSA would further have been interested in a method that allowed for the counting of the number of copies of a particular gene of interest in the original fragment. EX1011 (Iafrate) at 0029 (¶227). This is because cancer development was often associated with aberrant numbers of copies of particular oncogenes. EX1074 (Lonigro) at 0001–11. Thus, a cancer researcher would be interested in knowing whether there were

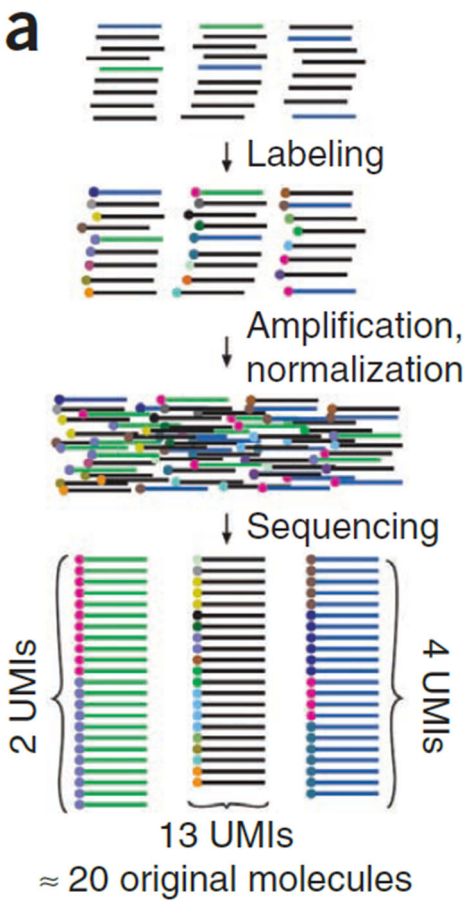
additional copies of a particular oncogene in a sample above the number expected in a normal human genome. But in order to perform this analysis, the researcher would need a means of distinguishing among the PCR copies of a particular sequence created during library preparation and SBS.

235. As explained above in Section IV.K.2.b, several publications that pre-date the '399 patent's priority date taught that using a molecular tag to differentiate nucleic acid fragments in NGS would mitigate PCR and sequencing error and improve the overall accuracy of NGS, including accurately counting the number of original fragments. EX1022 (McCloskey) at 0006; EX1026 (Casbon) at 0001–07; EX1028 (Jabara) at 0001–06; EX1030 (Shiroguchi) at 0001–05; EX1031 (Schmitt) at 0005. And several of those publications also taught the combination of a molecular tag and the sequence of the nucleic acid fragment could be used to identify duplicate sequencing reads. EX1024 (König) at 0007; EX1028 (Jabara) at 0002; EX1029 (Tollervey) at 0002; EX1030 (Shiroguchi) at 0005. Kivioja is one such publication.

236. Kivioja introduced a method to "improve the accuracy of almost any next-generation sequencing method." EX1012 (Kivioja) at 0001. Kivioja recognized the PCR errors create accuracy problems in NGS. EX1012 (Kivioja) at 0002. So Kivioja devised a method to solve the accuracy problem wherein a random molecular tag sequence that is added to an adaptor, in combination with the sequence

of the nucleic acid fragment to which that adaptor is appended, are used to associate observed sequences with a particular original fragment. EX1012 (Kivioja) at 0001, 0005. In other words, Kivioja taught a method of using a molecular tag and the associated nucleic acid sequence to identify duplicate sequencing reads.

237. Kivioja diagrammed the general method in Figure 1(a), which I have reproduced below. EX1012 (Kivioja) at 0001.



238. To accomplish the detection and removal of duplicate sequencing reads, as shown in Figure 1(a), Kivioja first fragmented a nucleic acid sample and

then labeled the fragments with unique molecular tags. EX1012 (Kivioja) at 0001. This "generate[d] a library in which each molecule has a distinct identifying sequence." EX1012 (Kivioja) at 0001. After amplification and sequencing, Kivioja taught that duplicate reads with the same label (molecular tag) and nucleic acid fragment sequence, *i.e.* reads with a unique label-fragment combination, corresponded to a single "UMI" or unique original sequence. EX1012 (Kivioja) at 0001. Indeed, Kivioja explains that a library of labeled fragments could "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." EX1012 (Kivioja) at 0001.

239. Kivioja then explained that "[u]pon 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." EX1012 (Kivioja) at 0001. Kivioja also showed this process in Figure 1(a), where the authors collapsed the labeled and amplified nucleic acid fragments to 20 original molecules. EX1012 (Kivioja) at 0001. That is, Kivioja identified duplicate molecules that were the result of PCR amplification and associated them to a particular original labeled fragment. EX1012 (Kivioja) at 0001.

240. Thus, in Figure 1(a), Kivioja disclosed generally the steps of claim 1 of

the ’399 patent. Kivioja (a) attached an adaptor with a molecular tag to the 5′ end of a DNA fragment, (b) subjected the labeled fragments to PCR amplification, (c) sequenced those amplification products, and (d) identified duplicate sequencing reads based on the UMI—that is, the unique combination of the molecular tag sequences and fragment sequences. EX1012 (Kivioja) at 0001, 0004–05.

241. Kivioja also conducted and reported several proof of concept experiments. In one experiment Kivioja started with mRNA molecules from a *Drosophila* fly. EX1012 (Kivioja) at 0002-05. Because Kivioja started with mRNA, and Illumina sequencers read DNA, Kivioja needed to convert the mRNA to cDNA. EX1012 (Kivioja) at 0003. To accomplish this, Kivioja used a protocol called the Switching Mechanism at the 5′ end of RNA Template (SMART). EX1012 (Kivioja) at 0004. Using this protocol, Kivioja added an adaptor to the 5′ end of the target fragments during the processes of reverse transcription wherein the mRNA was converted to cDNA using a certain reverse transcriptase. EX1012 (Kivioja) at 0004.

242. The adaptors Kivioja added to the 5′ ends of the cDNA fragments contained the same Illumina universal target sequence primer binding site found in Iafrate (in **green** below), a random molecular tag (in **yellow** and represented by the capitalized “N”s), and an indexing site (in **orange**), as shown in the annotated version of Kivioja’s Supplementary Table 1 that I have reproduced below. EX1012 (Kivioja) at 0002, 0014.

Supplementary Table 1. Sequences for oligomers used in mRNA-seq. Sequences in square brackets and parentheses represent an index sequence that was used to enable multiplexed sequencing, and a random ten base DNA sequence label, respectively. dU and rG represent deoxyuridine and guanine ribonucleotide, respectively.

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

243. Though Kivioja did not label a sample index primer binding site in the above-described adaptor, Kivioja's adaptor still contained a sample index primer binding site. The sample index, molecular tag, and fragment were all sequenced together on the Illumina device. EX1012 (Kivioja) at 0004–05. That means the green sequence in the Kivioja adaptor served as both the target sequence primer binding site and the index primer binding site. Thus, Kivioja disclosed an adaptor that comprised the same features as those recited in the '399 patent. EX1003 ('399 patent) at 25:42–46.

244. Though Kivioja used the template switch protocol to attach the above-described adaptor, the POSA would have understood that the template switch protocol could only be used with RNA. Kivioja also recognized that the template switch method was limited to RNA, as Kivioja described the method as one used to convert RNA to cDNA. EX1012 (Kivioja) at 0003.

245. But Kivioja's method, including using the elements of the adaptor, would have been applicable to DNA fragments. In Kivioja's other proof of concept

experiment, Kivioja disclosed that adaptors could be ligated to DNA fragments. EX1012 (Kivioja) at 0004. In that experiment, Kivioja pooled DNA from two patients and explained "[s]amples were prepared by enzymatic fragmentation, adaptor ligation and PCR" with standard Illumina Y adaptors. EX1012 (Kivioja) at 0004. In that experiment, Kivioja aliquoted the samples and used the fragment sequence alone ("the 5′ genomic position of each mapped read") as the UMI. EX1012 (Kivioja) at 0004. But the POSA would have recognized that ligated adaptors could be modified for use with genomic DNA if aliquoting was undesirable or impossible (e.g., because the target sequences of interest were at a low concentration).

246. Kivioja identified exactly this issue and noted that sequence labels would be required if a target sequence were at a low concentration. Kivioja explained, citing to Figure 1(a), that "[g]enerating UMIs for both high- and low-abundance species in a single reaction requires the use of sequence labels." EX1012 (Kivioja) at 0002. Figure 1(a) teaches labeling "DNA species," which could have been accomplished by ligating adaptors, as Kivioja recognized and as the POSA would have understood. Kivioja thus contemplated that adaptors containing molecular tags could be added to the 5′ end of DNA fragments using a ligation reaction. And it would have been obvious to the POSA to ligate to DNA fragments an adaptor containing the molecular tag disclosed in Kivioja, particularly for library

preparation and NGS methods that targeted low-abundance sequences and/or that used anchored PCR to identify unknown sequences (or both, as was the case with Iafrate's method).

247. Like Iafrate, Kivioja also taught the amplification and sequencing steps recited in claims 1(b) and 1(c) of the '399 patent. EX1012 (Kivioja) at 0004. Specially, with respect to one of Kivioja's mRNA experiments, the adaptor-nucleic acid fragments "were amplified using Phusion High-Fidelity DNA polymerase," which was primed with "Illumina single-read sequencing library primers." EX1012 (Kivioja) at 0004. The "PCR products were purified . . . and subjected to Illumina GAIIx massively parallel sequencing according to manufacturer's instructions." EX1012 (Kivioja) at 0004. For Kivioja's other mRNA experiment the adaptor-nucleic acid products were similarly amplified and then sequenced on an Illumina HiSeq 2000 instrument. EX1012 (Kivioja) at 0004.

248. Kivioja also taught the step recited in claim 1(d) of the '399 patent. Claim 1(d) requires "identifying a duplicate sequencing read comprising the same identifier site and nucleic acid fragment as another sequencing read in the population of sequencing reads." EX1003 ('399 patent) at 25:53–56. Kivioja explained that after sequencing, "the mapped reads with the same gene, position, and label were collected to one UMI and the number of such reads was recorded as the copy number of that UMI." EX1012 (Kivioja) at 0005. Specifically, for each gene, Kivioja

mapped the sequencing read against the "longest transcript" of that gene, which was used as the reference sequence. EX1012 (Kivioja) at 0005. Kivioja then discarded reads that, for example, did not map against the reference. EX1012 (Kivioja) at 0005. Then, for each gene, reads with the same "position," relative to the reference, and same "label" (molecular tag) were collapsed into a single UMI. EX1012 (Kivioja) at 0005. In sum, Kivioja described a method of using the molecular tag and sequence of the nucleic acid fragment to determine whether particular sequence reads were unique or PCR duplicates. Kivioja then deduplicated those reads by collapsing them into the single, unique UMI.

249. The POSA would have been motivated to combine Iafrate and Kivioja to practice the claims of the '399 patent. Iafrate taught a method of target enrichment that promised to enhance understanding of genetic sequences that influence cancer onset and progression, and that might have served as therapeutic targets. But the POSA would have recognized that Iafrate's method required multiple rounds of PCR followed by NGS sequencing, and that PCR-introduced error and sequencing error could hinder the method's capability of accurately determining the sequences adjacent to known cancer genes or accurately counting original molecules, such as copy number variants. The POSA would thus have sought to mitigate PCR errors and improve the accuracy of Iafrate's method. Kivioja taught that PCR errors could be mitigated by adding a random molecular tag sequence to an adaptor and using

that sequence, in combination with the sequence of the nucleic acid fragment, to detect and remove PCR errors. Both Iafrate and Kivioja moreover taught that their methods were suitable for use in NGS on Illumina platforms. Thus, the POSA would have been motivated to modify Iafrate's method and incorporate Kivioja's teachings.

5. The POSA Would Have Had a Reasonable Expectation of Success in Combining Iafrate and Kivioja to Arrive At The Claimed Method.

250. The POSA would have understood that Iafrate's method was "adaptable for any [] [NGS] platform." EX1011 (Iafrate) at 0029. Both Iafrate and Kivioja taught the use of Illumina NGS platforms. Given the overlapping teachings related to use on Illumina NGS platforms, and indeed their preference for Illumina platforms, the POSA would have had confidence combining the two methods. In fact, the Iafrate adaptor and the Kivioja adaptor incorporated the same universal Illumina sequence, as shown below in **green**. Based on similarity between the adaptors disclosed in Kivioja and Iafrate, and the POSA's understanding that NGS methods involved generally interchangeable adaptor sequences, the POSA would have readily understood that Kivioja's molecular tag could be inserted into the analogous position in the Iafrate adaptor. That is, the POSA could have inserted the random molecular tag on the 3′ side of the sequencing primer binding site, which I have annotated with a star below in my annotated comparison of Iafrate Figure 9 and Kivioja Supplemental Table 1, when designing the adaptor sequence. EX1011

(Iafrate) at 0011; EX1012 (Kivioja) at 0014.

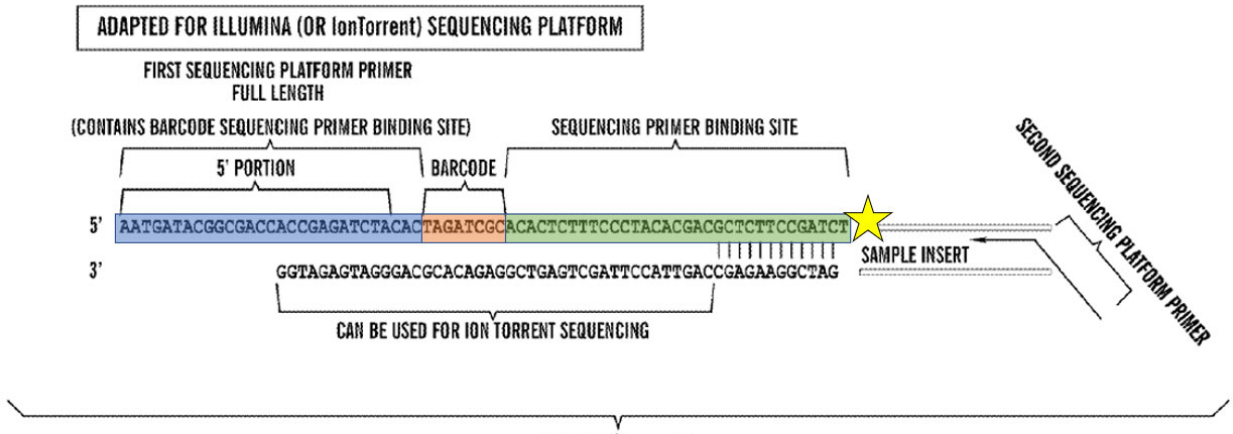


FIG. 9

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

251. The POSA further would have understood that the relevant portions of the adaptor could be easily moved or redesigned as needed by a particular experimental design. For example, the POSA could have inserted the identifier site into the Iafrate adaptor before or after the barcode/sample index (orange), so as to read the identifier site as part of the sample index read. Or the POSA could have moved the barcode site in the Iafrate adaptor to the 3′ side of the sequencing primer binding site, e.g., to read the sample index and the molecular tag together with the fragment sequence in one read, or as part of a paired end read with a short forward read through the barcode and the identifier, and then a comparatively long read through the fragment in the other direction in the complementary strand. All of these

permutations were well known design choices that were routine in the design of Illumina adaptors at the priority date.

252. After designing the sequences of the adaptors as described above, it would have been routine at the priority date for the POSA to purchase the relevant adaptor sequences from a commercial vendor, as both Iafrate and Kivioja did. EX1011 (Iafrate) at 0019 (¶81); EX1012 (Kivioja) at 0004. In the late 2000s and early 2010s, I routinely ordered synthetic oligonucleotides for use in NGS library preparations. Thus, it would have been trivial for the POSA to incorporate a molecular tag into the Iafrate adaptors.

253. Given the overlap between the adaptors used in both Iafrate and Kivioja and the similarity in the other steps of library preparation, the POSA would have had more than a reasonable expectation of success that the two methods could be combined.

6. Claims 1–3, 5–11, and 14–17 Of The '399 Patent Would Have Been Obvious Over Iafrate and Kivioja.

a. Claim 1

254. Claim 1 is reproduced below.

1. 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.

255. Claim 1 would have been obvious to the POSA at the priority date based on Iafrate and Kivioja. The POSA would have been motivated to use Iafrate's method of library preparation where target enrichment prior to sequencing was desired, e.g., for cancer research. The POSA further would have been motivated to improve Iafrate's method, which would have been hampered by PCR and sequencing error, by incorporating Kivioja's molecular tags and use of UMIs. Based on this combination, the POSA would have been motivated to practice a method comprising all of the recited steps of claim 1 and use an adaptor with all of the

claimed features, and would have had a reasonable expectation of success in doing so.

1) Claim 1: "A method for detecting a duplicate sequence read from a population of sample sequencing reads"

256. Specifically, claim 1 of the '399 patent recites "a method for detecting a duplicate sequencing read from a population of sample sequencing reads." EX1003 ('399 patent) at 25:37. As I discussed above in Sections VI.C.3, the method recited in claim 1 requires an adaptor with features that Iafrate disclosed. The only feature Iafrate did not disclose was the molecular tag, but, as I discussed above in Sections VI.C.4-5, Kivioja did and the POSA would have been motivated to modify the Iafrate method to incorporate Kivioja's teachings. Kivioja taught a method that included, among other things, detecting a duplicate sequencing read from a population of sample sequencing reads. As I discuss in more detail above in ¶¶ 238-239 of my declaration, Kivioja identified duplicate sequencing reads as those with the same molecular tag sequence and fragment sequence. EX1012 (Kivioja) at 0005.

257. Kivioja also taught that one of the principal benefits of the method was for digital karyotyping. EX1012 (Kivioja) at 0001–02. Karyotyping is a genetic test that, among other things, can be used to determine the number of chromosomes or copies of a particular gene in a cell to be counted, a process that is typically conducted with fluorescence-aided microscopy. A karyotype is commonly used as

a method of diagnosis for certain disorders and diseases.

258. Kivioja's UMI method allows for the identification of PCR copies of original genetic fragments. EX1012 (Kivioja) at 0001–02. Because the PCR copies of original fragments can be identified, the UMI method enables reliably counting the absolute number of molecules in a sample. EX1012 (Kivioja) at 0001–02. Kivioja reported the accuracy of the UMI method applied to karyotyping to be "close to the theoretical limit imposed by the sample size, suggesting the development of the UMI method for diagnostic use appears feasible." EX1012 (Kivioja) at 0002.

259. Kivioja then applies the UMI method to mRNA sequencing to demonstrate that "one can reproducibly detect the number of molecules after different numbers of PCR cycles without normalization." EX1012 (Kivioja) at 0002–03. This was not the case prior to applying the UMI method, when "after 15 or 25 PCR cycles," Kivioja recorded "amplification bias that resulted in loss of accuracy." EX1012 (Kivioja) at 0002. Thus, the UMI method was "likely to improve a large number of next-generation sequencing-based molecule-counting applications and also enable new methods for tracking relationships between molecules." EX1012 (Kivioja) at 0002.

260. Kivioja's method of counting the absolute number of molecules, whether in its karyotyping experiment or either of its mRNA experiments, necessarily requires the identification of PCR duplicates. The POSA would have

understood that, using Kivioja's method, the number of molecules present in an original sample could not be counted unless UMIs were used to first determine whether a particular observed sequence is a PCR duplicate of an original fragment. Otherwise PCR duplicates would be counted as original molecules, thus resulting in an inaccurate molecular count. While Kivioja does not use the word "duplicate," the POSA would have recognized that Kivioja teaches a method of detecting and removing duplicate sequencing reads.

2) Claim 1: Step a)

261. Claim 1 next recites "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." EX1003 ('399 patent) at 25:39–46. For the reasons I discuss above in Sections VI.C.3–5 of this declaration, the combination of Iafrate and Kivioja disclosed this limitation. Both Kivioja and Iafrate taught ligation of adaptors to the 5′ end of nucleic acid fragments prior to NGS sequencing. And both references taught adaptors containing indexing primer binding sites, indexing sites, and a target sequence primer binding site, specifically the standard Illumina sequences that served those functions in adaptors for use on the prior art Illumina NGS devices. Kivioja taught an identifier site, and as I explain above in Section

VI.C.4 of this declaration, the POSA would have been motivated to incorporate this identifier site into Iafrate's adaptor.

262. Claim 1 also requires the identifier site to be "between 1 and 8 nucleotides." EX1003 ('399 patent) at 25:44–45. Modifying the Kivioja method to include a molecular tag in this range of length would have been obvious. While Kivioja taught a molecular tag comprised of 10 random bases, the POSA would have recognized that molecular tags of fewer or more nucleotides could be used depending on the particular application.

263. As I explain above in Section IV.K.2.b of my declaration, at the priority date, molecular tags of varying lengths were disclosed in the prior art, and the POSA would have understood that it would have been beneficial to customize the length of the molecular tag depending on certain experimental conditions. When the molecular tag sequence was read along with the fragment sequence, as in Kivioja, the prior art taught that, because total sequencing read length was limited on all of the various NGS platforms, the tag should be kept to a minimum so as to maximize the amount of the fragment sequence included in the sequencing read. This was particularly true when sequencing on an Illumina device in 2013, given the especially limited read length of the most common Illumina models relative to other platforms such as the 454. Accordingly, the POSA would have wanted to keep the molecular tag length as short as possible while ensuring a sufficient number of

molecular tags to make each fragment unique.

264. The POSA would have recognized, as taught in McCloskey and other prior art references, which I discussed in Section IV.K.2.b of my declaration, that the number of unique molecular tags was a simple function of the length of the molecular tag (4^n), EX1022 (McCloskey) at 0004; EX1026 (Casbon) at 0001, 0005. The POSA would further understand that there must be an excess of tags relative to the number of unique target molecules in order to render the probability of getting "doublets" (that is, a single unique tag associated with more than one unique molecule) negligible. As of the priority date, the relationship between the number of unique tags needed to ensure efficient and unique labeling of a given number of target molecules was well-understood in the art, as shown in the figure I have reproduced below from EX1027 (Fu) at 0002. Because the "expected number of unique labels captured depends on the ratio of molecules to labels," EX1027 (Fu) at 0002, if the number of target molecules changes, so too does the optimal number of labels required to ensure efficient and unique labeling.

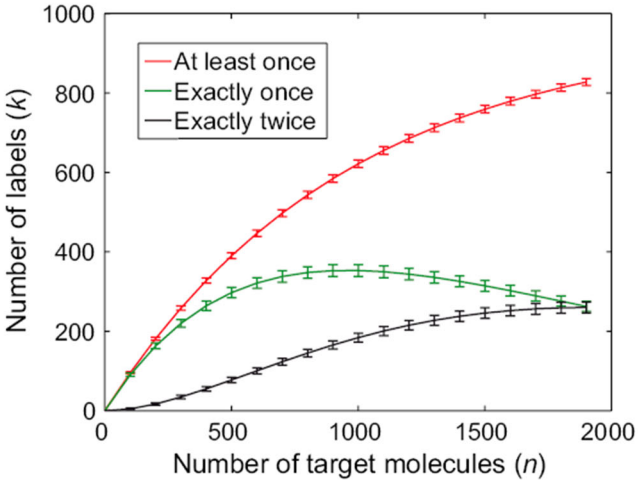


Fig. 2. The number of stochastically captured labels for a given number of target molecules calculated using a nondepleting reservoir of 960 diverse labels. The red curve represents the average number of labels observed at least once (calculated from Eq. **S1**); the green and black curves represent the number of labels observed exactly once and twice (calculated from Eq. **S3**), respectively. Error bars indicate one standard deviation (calculated from Eqs. **S2** and **S4**) away from the corresponding mean values.

265. Consistent with Fu (EX10127) and other prior art, the POSA would further understand that, for sequencing applications involving a single target sequence, the number of unique fragments in a sample can be estimated based on the length distribution of the fragments (which can be estimated according to the method of fragmentation or empirically measured), the mass of the sample, and, if applicable, a correction factor that accounts for the amount of PCR amplification performed to obtain the sample. For shotgun (e.g. genome- or transcriptome-wide sequencing) applications, one must estimate the copy number of the most abundant unique fragments in the sample to determine the appropriate number of molecular tags. The same calculation described above can be used to estimate the number of

molecules in the original sample, and the copy number of the most abundant molecules can then be estimated from prior gene expression or copy number profiles for the same or similar samples. More precise estimates can be obtained using methods like digital PCR. I have done this calculation myself on countless occasions in order to design molecular tagging experiments.

266. In this regard, I note that the length of the molecular tag is a "result-effective variable" (as I understand that term is defined in the law set out in Section III.D, ¶ 27 above)—where the length of the tag has a predictable relationship with the total number of unique tags, which in turn has a predictable relationship with the expected number of unique fragments of interest in a particular sample. The calculations underlying these relationships were routine and well-known as of the priority date, and are ones that I have routinely undertaken when designing experiments in my own research. The POSA thus would have been motivated to decrease the length of the molecular tag if significantly fewer than 4^n molecular tags were required to minimize the number of doublets given the expected number of unique fragments.

267. For applications like Iafrate, where only some sequences of interest were being targeted, the POSA would have recognized that there would have been relatively few fragments containing the relevant sequences of interest in a low-input sample from relatively few cells or a small piece of tissue. In such low-input

situations, compared to a situation where a high-input sample containing a large number of cells or a large piece of tissue, the POSA would have recognized that the molecular tag could be fewer than 10 nucleotides, and it would have been obvious to modify the method to use a molecular tag in the range of 1 to 8 nucleotides.

268. For example, although Kivioja incorporated 10 nucleotides into the molecular tag, Kivioja also explained that using 10 nucleotides yielded "approximately one million random labels" that "were not exhausted in the experiments." EX1012 (Kivioja) at 0005. Kivioja thus had a greater number of molecular tags than unique nucleic acid fragments in the sample. Given this, the POSA would have understood that the length of the molecular tag could be varied depending on, for example, the number of nucleic acid fragments in the original sample.

269. The POSA would have been motivated to apply this teaching to Iafrate's method. Iafrate explained that in certain applications of its method the number of fragments of interest would be relatively low, including for the detection of mutations associated with particular cancer genes (for example in Iafrate's proof of concept, where the inventors looked at only three genes, ROS1, ALK, and RET, in a particular NGS run). EX1011 (Iafrate) at 0029. In addition, Iafrate also explained that the number of PCR cycles could be optimized based, in part, "on how much starting nucleic acid material is used." EX1011 (Iafrate) at 0029. The POSA

would have known that for a lower concentration sample, more PCR cycles might be needed. The POSA would have understood that for the targeted enrichment methods of Iafrate, particularly when combined with a relatively low sample concentration, the ten nucleotide molecular tag used in Kivioja, which was excessive even for Kivioja's shotgun sequencing method, would have been optimized to a lower number, including a molecular tag with between 1 and 8 nucleotides.

270. Furthermore, I note that other prior art groups used an identifier site in the range of 1 to 8 nucleotides, including Miner, EX1021 (Miner) at 0002 (7 nucleotides), König, EX1024 (König) at 0007 (3 nucleotides), McCloskey, EX1022 (McCloskey) at 0002 (7 nucleotides), and Jabara, EX1028 (Jabara) at 0002 (8 nucleotides). This further demonstrates that the length of the identifier site was the product of routine optimization for the needs of a particular experiment, as I explain above in ¶¶ 264-266.

271. In addition, I do not understand the '399 patent to disclose any particular benefit to the range of 1 to 8 nucleotides. The patent does not disclose any unexpected properties or results of that range of nucleotides, let alone compared to a molecular tag of 10 nucleotides as used in Kivioja. Indeed, the '399 patent does not include any relevant data to support the claimed length of identifier site. To the contrary, the prior art taught that the selection of a particular length of molecular tag was largely arbitrary so long as there were more molecular tags than expected unique

fragments, as explained above. Accordingly, the POSA would routinely optimize the length of the molecular tag for a particular application and thereby arrive at a molecular tag of between 1 and 8 nucleotides for applications where an appropriate number of unique fragments were expected.

3) Claim 1: Steps b) and c)

272. Claim 1 next recites "b) amplifying the adapter-nucleic acid fragment ligated products" and "c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated fragments."

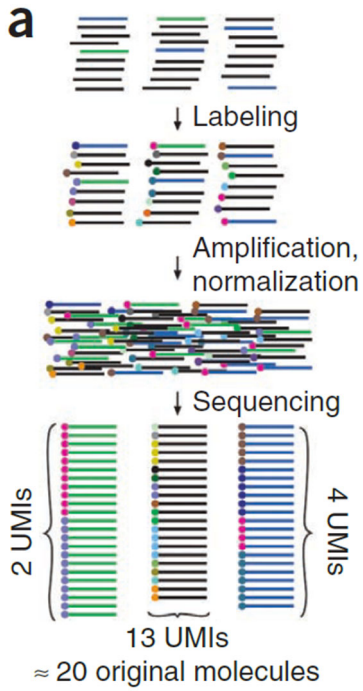
273. As I discuss above in Sections VI.C.3–4 of this declaration, Iafrate and Kivioja teach these steps as part of their methods. Specifically, both Iafrate and Kivioja amplified the adapter-fragment products prior to sequencing. EX1011 (Iafrate) at 0012–13 (¶¶7,14), 0017–20 (¶¶64-94), 0051 (claim 1(b) and 1(c)); EX1012 (Kivioja) at 0001 (Figure 1(a)), 0003–05. And Iafrate and Kivioja both taught sequencing of those fragments on an NGS device to arrive at a population of sequencing reads. EX1011 (Iafrate) at 0012 (¶7), 0021 (¶100), 0029 (¶232), 0051–52 (claims 1 and 11).

4) Claim 1: Step d)

274. Claim 1 also requires "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." EX1011 (Kivioja)

at 0021. Kivioja discloses this limitation for the reasons I discuss above in Sections VI.C.4 and VI.C.6.a.1 of my declaration, as well as the reasons explained in this section.

275. First, in Figure 1(a), Kivioja shows that sequences are collapsed to UMIs based on the colored label sequence (the molecular tag) and the colored fragment sequence. EX1012 (Kivioja) at 0001. This is shown best by the pink molecular tag. The pink molecular tag is associated with the green DNA sequence fragments and the blue DNA sequence fragments. Yet Kivioja shows that the pink + green UMI and the pink + blue UMI were recorded separately. If duplicate sequencing reads were identified only by the molecular tag, this sorting would not have been possible.



276. Second, in its mRNA example, Kivioja describes identifying a duplicate sequencing read as a sequencing read "with the same gene, position and label." EX1012 (Kivioja) at 0005. As described, for each "gene" of interest, to identify a read with the same "position," Kivioja mapped the fragment sequence read to a reference gene from the *Drosophila* genome. EX1012 (Kivioja) at 0005. The POSA would have understood this to mean that Kivioja used the "nucleic acid fragment" to help determine a duplicate sequencing read. Kivioja combined the mapped position of the nucleic acid fragment with the label sequence to fully assess whether sequencing reads were PCR duplicates of an original fragment sequence. EX1012 (Kivioja) at 0005. So put together, Kivioja discloses identifying a duplicate sequencing read as a sequencing read comprising the same identifier site (the "label") and nucleic acid fragment (the "gene" and "position") as another sequencing read in the population of sequencing reads.

277. Third, in the karyotyping proof of concept experiment, Kivioja explained that, after sequencing, the "reads were mapped to the genome." EX1012 (Kivioja) at 0004. Mapping reads to the genome necessarily required matching the observed sequence of the nucleic acid fragment to the expected sequence. Kivioja then explains that, "for absolute molecule counting," "the 5′ genomic position of each mapped read" defined the UMI. EX1012 (Kivioja) at 0004. Using the sequence of the nucleic acid fragment in this way—first mapping the entire fragment

sequence to the genome and then using the 5′ position—Kivioja was able to verify "that UMIs indeed represented single molecules." EX1012 (Kivioja) at 0004.

278. Fourth, Kivioja explained that the molecular tag sequence "and the 5′ mapped position of the fragment together define the UMI." EX1012 (Kivioja) at 0003. To locate the 5′ mapped position Kivioja first mapped the sequencing reads to the reference genome. EX1012 (Kivioja) at 0005. Kivioja then explained that "read counting biases introduced during PCR or *in silico* could be identified using the UMI method." EX1012 (Kivioja) at 0003. Because Kivioja first mapped the sequence of the nucleic acid fragments to a reference genome and then used "the 5′ mapped position of the fragment" to identify PCR biases, or PCR-introduced error, Kivioja necessarily used the sequence of the nucleic acid fragment (and molecular tag) to identify PCR duplicates.

5) Claim 1: "wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides"

279. Claim 1 further recites that "the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides." EX1003 ('399 patent) at 25:56–58. Both Iafrate and Kivioja teach this limitation, which is a standard implementation of sample indexes at the priority date. As explained above in Sections VI.C.3–4 of my declaration above, both Iafrate

and Kivioja teach the use of sample indexes for multiplexing. Sample indexes are unique among a subset of nucleic acid fragments from a particular sample and serve as an index to those fragments.

280. Iafrate teaches that "the methods described herein also allow for multiplex sample pooling after the Y adapter ligation during which individual samples are ligated with unique barcoded adapters." EX1011 (Iafrate) at 0029 (¶231). Iafrate goes on, "[o]nce barcoded via Y adapter ligation, multiple samples may be pooled into on reaction tube for downstream target enrichment." EX1011 (Iafrate) at 0029 (¶231). As Iafrate explains "[e]ach resulting sequencing read of an amplification product will therefor comprise a barcode identifying which sample comprised the original template nucleic acid from which the amplification product is derived." EX1011 (Iafrate) at 0023 (¶112).

281. Kivioja likewise explains that its 5′ adapter included a sample index "that was used to enable multiplexed sequencing." EX1012 (Kivioja) at 0014.

6) Claim 1: "wherein the sequence of the identifier site is variable in sequence content in the plurality of adapters"

282. Claim 1 also recites that "the sequence of the identifier site is variable in sequence content in the plurality of adapters." EX1003 ('399 patent) at 25:59–61. Kivioja discloses this limitation. Kivioja's molecular tag is "a random ten base DNA sequence label." EX1012 (Kivioja) at 0014. In Kivioja's Supplementary

Table 1, the 10 random bases that make up the molecular tag are identified as a series of "N"s. EX1012 (Kivioja) at 0014. The POSA would have understood that this was a common notation for random bases. Because Kivioja used random bases for the molecular tags, which were incorporated into adaptors, and which the POSA would have been motivated to incorporate in Kivioja's adaptors, the molecular tags would have been variable in sequence content. Moreover, the POSA would have known that the prior art generally taught that it would have been beneficial to use random (variable) molecular tags for differentiating between original nucleic acid fragments and PCR copies.

7) Claim 1 Would Have Been Obvious Over Iafrate and Kivioja

283. For the reasons stated above in Sections VI.C.6.a.1–6, claim 1 would have been obvious to the POSA at the priority date based on Iafrate and Kivioja. The POSA would have been motivated to use Iafrate's method of library preparation and would have been motivated to improve Iafrate's method by incorporating Kivioja's molecular tags and use of UMIs. Based on this combination, the POSA would have been motivated to practice a method comprising all of the recited steps of claim 1 and use an adaptor with all of the claimed features, and would have had a reasonable expectation of success in doing so.

b. Claim 2: "The method of claim 1, wherein the method further comprises removing from the population of

sequencing reads the sequencing read with a duplicate identifier site and target sequence."

284. Claim 1 would have been obvious to the POSA for the reasons explained above in Sections VI.C.6.a.1-7.

285. Claim 2 would have been obvious based on Iafrate and Kivioja. As explained above in Sections VI.C.4, VI.C.6.a.1, and VI.C.6.a.4 of my declaration, Kivioja teaches that duplicate sequencing reads are identified as having the same molecular tag sequence and nucleic acid fragment sequence. And Kivioja teaches that duplicate reads are "collected to one UMI" and treated as a single original fragment—necessarily removing the duplicates from the population of sequencing reads. EX1012 (Kivioja) at 0005, 0001. Also as described above in Section VI.C.6.a.1, Kivioja reported a method of counting the original number of molecules in a sample, which the POSA would have understood necessarily requires deduplication, and removing of the duplicate reads, as a predicate step.

286. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 2.

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the indexing site."

287. Claim 1 would have been obvious to the POSA for the reasons explained above.

288. The ordinary meaning to the POSA of sequencing the identifier site

with the indexing site is that both sequences are sequenced in a single SBS read on the NGS device. Under that interpretation, claim 3 would have been obvious based on Iafrate and Kivioja. Kivioja constructed its 5′ adaptor with the sample index and molecular tag on the 3′ end of the of the sequencing primer binding site. EX1012 (Kivioja) at 0014. And as I explained above in Sections VI.C.4–5, the POSA would have been motivated to modify Iafrate's method to add the molecular tag taught by Kivioja to the 3′ side of the sequencing primer binding site in Iafrate's adaptor. The placement of the sequencing primer binding site relative to the sample index and molecular tag means that the sample index and molecular tag would have been sequenced in a single SBS read.

289. For Kivioja's mRNA experiment, the adaptor-cDNA fragments were sequenced in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at 0004. The POSA would have understood Kivioja's disclosure of 54-bp reads to mean the identifier site was sequenced with the sample index because Kivioja discloses only a single read of the adaptor-cDNA fragment.

The POSA's understanding would have been confirmed by Kivioja's separate disclosures of paired-end sequencing in the genomic DNA experiment and other mRNA experiment. For the gDNA experiment, Kivioja explained that "sequences were generated on an Illumina Genome Analyzer with 76-bp single reads for the mixed sample and 100-bp paired-end reads for the adult male sample." EX1012

(Kivioja) at 0004. In the other mRNA experiment, Kivioja explained that the samples were sequenced on an Illumina HiSeq 2000 with "54-nucleotide reads from both ends." EX1012 (Kivioja) at 0004. With the molecular tag and sample index in Kivioja's method being on the 5′ side of the nucleic acid fragment, and not on the 3′ side, paired-end reads would have resulted in the identifier site and sample index being sequenced together in one direction, but not in the other. Kivioja, moreover, distinguished between the use of single read sequencing and paired-end sequencing, as in the DNA experiment. EX1012 (Kivioja) at 0004. The POSA would thus have understood Kivioja to teach that the molecular tag and sample index in one mRNA experiment were sequenced together in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at 0004.

290. To the extent that claim 3 requires the identifier site and the indexing site to be placed in the adaptor in such a position as to permit them to be sequenced together in a single SBS read, the choice of where to place the sample index and molecular tag in any particular adaptor would have been a design choice for the POSA. Depending on the NGS platform, the sample index and molecular tag would necessarily be to the 3′ side of the sequencing primer; for example, if 454 were used. For other devices, such as certain Illumina devices, the sample index could be read from a dedicated sample index primer binding site while the molecular tag could be read from the sequence primer binding site. Thus, the POSA would have known

that the sample index and molecular tag could be placed in different positions within an adaptor depending on the NGS platform used or the requirements of an experiment.

291. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 3.

292. I further do not understand the '399 patent to disclose any unexpected properties or results of sequencing the indexing site and the identifier site in a single SBS read. To the contrary, the POSA would understand that there would be no practical difference, from the perspective of the POSA obtaining sequencing data, of sequencing the indexing site and the identifier site in two different SBS reads as part of a single run of the NGS device, versus sequencing them in a single SBS read.

d. Claim 5: "The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample."

293. Claim 1 would have been obvious to the POSA for the reasons explained above in Sections VI.C.6.a.1-7.

294. Claim 5 would have been obvious based on Iafrate and Kivioja. Iafrate and Kivioja both disclose methods that are applicable to multiplex sequencing. Iafrate discloses an adaptor with sample indexes, which are used for multiplex sequencing. EX1011 (Iafrate) at 0011. Iafrate also explains, for example, that the method "allow[s] for multiplexed sample pooling after the Y adapter

ligation . . . during which individual samples are ligated with unique barcoded adapters." EX1011 (Iafrate) at 0029. Some of Iafrate's claims also recite the use of sample indexes. EX1011 (Iafrate) at 0052.

295. Kivioja similarly discloses that its method is applicable to multiplex sequencing. As described, Kivioja discloses an adaptor with a sample index, which "was used to enable multiplexed sequencing." EX1012 (Kivioja) at 0014. Kivioja further explained that "[t]he UMI method is compatible with sample indexing using separate DNA barcodes, allowing parallel analysis of samples." EX1012 (Kivioja) at 0003. Kivioja also conducted several proof of concept experiments where nucleic acid fragments from multiple samples were sequenced together. EX1012 (Kivioja) at 0004–05.

296. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 5.

e. Claim 6: "The method of claim 5, wherein the nucleic acid fragments from each sample have the same indexing site and each sample has a unique indexing site."

297. Claims 1 and 5 would have been obvious to the POSA for the reasons explained above.

298. Claim 6 would have been obvious based on Iafrate and Kivioja. As just described, both Iafrate and Kivioja disclose the use of sample indexes, which are

used for pooling and sequencing multiple samples. Iafrate explains that "multiple samples" can be appended to adaptors, each "with a unique barcode portion and wherein the samples are pooled." EX1011 (Iafrate) at 0013. Kivioja similarly explains that "[t]he UMI method is compatible with sample indexing using separate DNA barcodes." EX1012 (Kivioja) at 0003. That both Iafrate and Kivioja explain that the sample indexes are "unique" or "separate" shows that Iafrate and Kivioja both disclose that each sample has its own unique indexing site and so each nucleic acid fragment from a particular sample has the same indexing site.

299. The POSA would also have understood that for multiplex sequencing to work it would have been necessary to ensure that the nucleic acid fragments from each sample had the same sample index and that each sample had a unique sample index, otherwise sequencing reads from multiplexed sequencing would not have been able to be associated back to their original sample. The POSA would also have understood that by 2013 this was consistent with the standard use of sample indexes for Illumina sequencing.

300. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 6.

f. Claim 7: "The method of claim 6, wherein the sequencing reads are separated based on the indexing site."

301. Claims 1, 5, and 6 would have been obvious to the POSA for the

reasons explained above.

302. Claim 7 would have been obvious based on Iafrate and Kivioja. As described above, Iafrate and Kivioja both describe the use of sample indexes for multiplex sequencing. Pooling samples for a single sequencing run would have required separating the sequencing reads based on the sample index.

303. The POSA would also have known using sample indexes for multiplexed sequencing, followed by separation of samples based on indexing site, was consistent with the standard Illumina process in 2013. The process of separating based on indexing site would have been referred to as "demultiplexing." EX1060 (Meyer & Kircher) at 0009; EX1061 (Kircher Thesis) at 0053; EX1062 (Smith) at 0034. After a multiplexed sequencing run, Illumina devices would have output sequencing data in FASTQ files that were separated by sample based on the sample index sequences. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35.

304. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 7.

g. Claim 8: "The method of claim 7, wherein the separation of sequencing reads is performed prior to step d)."

305. Claims 1, 5, 6, and 7 would have been obvious to the POSA for the reasons explained above.

306. Claim 8 would have been obvious based on Iafrate and Kivioja. As

described with respect to claim 1 above in Section VI.C.6.a.5 above, Iafrate and Kivioja both disclose the use of sample indexes suitable for use on Illumina devices. The POSA would have known that by 2013 the standard process for demultiplexing sequencing data from an Illumina device would have been to separate the samples based on the sequences found in the FASTQ file data from the index read(s) before doing any further data processing, EX1062 (Smith) at 0034–38; EX1063 (Won) at 0003–04, such as identifying a duplicate sequencing read, EX1054 (Loeb) at 0012–15.

307. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 8.

h. Claim 9: "The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments."

308. Claim 1 would have been obvious to the POSA for the reasons explained above.

309. Claim 9 would have been obvious based on Iafrate and Kivioja. . Iafrate teaches that the nucleic acid fragment can be "any molecule, preferably a polymeric molecule, incorporating units of ribonucleic acid, deoxyribonucleic acid or an analog thereof," including RNA and DNA. EX1011 (Iafrate) at 0015 (¶44). Iafrate's examples likewise used DNA. EX1011 (Iafrate) at 0028 (¶¶230–31, 233, 235, 240–47). Kivioja similarly teaches that its method is applicable to both DNA

and RNA. EX1012 (Kivioja) at 0001–03. Kivioja includes examples with both DNA and RNA. EX1012 (Kivioja) at 0002–05. Moreover, the POSA would have known that molecular tags could be used to identify duplicate sequencing reads derived from DNA or RNA samples.

310. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 9.

i. Claim 10: "The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."

311. Claims 1 and 9 would have been obvious to the POSA for the reasons explained above.

312. Claim 10 would have been obvious based on Iafrate and Kivioja. Iafrate teaches that the nucleic acid fragment can be "any molecule, preferably a polymeric molecule, incorporating units of ribonucleic acid, deoxyribonucleic acid or an analog thereof. . . . including genomic DNA and cDNA." EX1011 (Iafrate) at 0015 (¶44). For instance, Iafrate explains in Example 1 that its method "enables targeted resequencing of cDNA or gDNA for potential detection of genetic alterations (single nucleotide variants, insertions/deletions, and copy number), epigenetic alterations (methylation), gene expression, and genomic rearrangements." EX1011 (Iafrate) at 0029. Iafrate used genomic DNA in its examples. EX1011 (Iafrate) at 0028 (¶¶230, 240–47). Kivioja similarly taught that

its UMI method could be applied to genomic DNA and cDNA. EX1012 (Kivioja) at 0001–03. Kivioja used genomic DNA in one example and cDNA in its mRNA example. EX1012 (Kivioja) at 0002, 0004. Moreover, the POSA would have known that molecular tags could be used to identify duplicate sequencing reads derived from genomic DNA and cDNA samples.

313. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 10.

j. Claim 11: "The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length."

314. Claim 1 would have been obvious to the POSA for the reasons explained above.

315. Claim 11 would have been obvious based on Iafrate and Kivioja. Both Iafrate and Kivioja teach the use of sample indexes that are between 2 and 8 nucleotides in length. In Iafrate's adaptor disclosed in Figure 9, the sample index (referred to as a "barcode") is 8 nucleotides. EX1011 (Iafrate) at 0011. In Kivioja's adaptor, the sample index (referred to as an "index sequence") is 5 nucleotides. EX1012 (Kivioja) at 0014.

316. Furthermore, even if Iafrate and Kivioja did not teach indexing sites within the range of 2 to 8 nucleotides, it would have been obvious to the POSA to modify the method to arrive at a sample index in that range via routine optimization.

At the priority date, the POSA would have understood that sample indexes could be designed to different lengths depending on the design needs of a particular application. The POSA would have recognized that the length of the sample index was simply dependent on the number of samples and the length of sequence required to ensure to the POSA that samples would remain uniquely labeled even accounting for PCR and sequencing error. To multiplex more samples, the POSA could use longer sample indexes, and with fewer samples, the POSA could use shorter sample indexes. Based on this predictable relationship, the POSA would have arrived at a sample index between 2 and 8 nucleotides based on routine optimization.

317. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 11.

318. In addition, I do not understand the '399 patent to teach any unexpected results or benefits of the particular length of 2 to 8 nucleotides for an indexing site. To the contrary, this range of nucleotides is simply arbitrary and the POSA would not have recognized any special benefit to that length over, for example, a sample index of 9 nucleotides.

k. Claim 14: "The method of claim 1, wherein the identifier site is 8 nucleotides in length."

319. Claim 1 would have been obvious to the POSA for the reasons explained above.

320. Claim 14 would have been obvious based on Iafrate and Kivioja. As described above in Section VI.C.6.a.2, by the priority date, the POSA would have known that the molecular tag could be designed to be a number of different nucleotides in length, and that it would have been beneficial to alter the length depending on certain experimental conditions. As I discuss above, the POSA would have been motivated to routinely optimize the length of the identifier site to keep the length as short as possible (thus maximizing read length) while ensuring that the number of nucleotides was great enough to ensure more unique identifier sites (a number that could be trivially calculated for a random identifier site using the formula 4^n) than the expected number of unique fragments. Because of this, as I explain above in ¶¶ 264-266, the length of the identifier site is a "result-effective variable" (as I understand that term is defined in the law set out in Section III.D, ¶ 27 above) wherein there was a known relationship between the variable (length of the identifier site) and the parameter being optimized (the number of labels necessary to uniquely tag the number of unique original fragments expected). Based on this routine optimization, the POSA would have considered an identifier site of 8 nucleotides obvious.

321. Notably, as I discuss further above in Section VI.C.6.a.2, Kivioja's data demonstrated that the 10 nucleotide identifier site resulted in an excess of unique molecular tags compared to the number of unique fragments. Kivioja explained that

in the proof of concept experiment with a 10 nucleotide identifier site, "[t]he approximately one million random labels [4^{10} or 1,048,576] used were sufficient to generate UMIs from an mRNA amount that corresponds to the amount found in ~1000 *Drosophila* S2 cells." EX1012 (Kivioja) at 0005. Based on this amount of input mRNA, "[o]n average only 0.0005% of all labels were observed per position, and even in the position with the highest number of labels, less than 2% of over one million labels were observed." EX1012 (Kivioja) at 0005. Based on this result, the POSA would understand that the approximately one million tags were more than sufficient to tag the unique fragments (given that only 2% of the labels were reported at the most common fragment).

322. Kivioja further explained that "a large fraction of all labels (>70%) were observed, and less than 0.12% overlap of UMIs (position label pairs) was observed between replicate experiments, indicating that the UMIs were not exhausted in the experiments and that the labels were incorporated into cDNAs effectively randomly." EX1012 (Kivioja) at 0005. Based on this data, the POSA further would have had confidence that the 10 nucleotide identifier site used by Kivioja was longer than was needed to ensure that every unique fragment was uniquely tagged.

323. Beyond Kivioja itself, the prior art regarding molecular tags that I discuss above in Section IV.K.2.b made clear that the length of a random molecular

tag could be varied depending on the number of unique fragments expected. This is demonstrated by the fact that prior art groups used molecular tags that varied in length in the single-digit range. For example,

- König used a molecular tag of 3 nucleotides, EX1024 (König) at 0007,
- McCloskey used a molecular tag of 7 nucleotides, EX1022 (McCloskey) at 0003,
- Miner used a molecular tag of 7 nucleotides, EX1021 (Miner) at 0002, and
- Jabara used a molecular tag of 8 nucleotides, EX1028 (Jabara) at 0002.

324. Furthermore, work that I participated in that was published in Shiroguchi demonstrated that using far fewer unique molecular tags, approximately 21,000, was sufficient to uniquely label fragments from an *E. coli* cDNA library. EX1030 (Shiroguchi) at 0002. While the method we described in Shiroguchi used two unique designed barcodes ligated to either end of the fragment instead of random identifiers, the data we generated was consistent with the POSA's understanding at the priority date that for most applications, the number of unique tags generated by Kivioja's 10 nucleotide identifier site was in excess of the number of unique fragments.

325. Given this, the POSA would have understood that the length of the molecular tag could be varied depending on, for example, the number of nucleic acid fragments in the original sample, and that 10 bases resulted in more unique

molecular tags than there were unique fragments in the transcriptome shotgun sequencing method Kivioja tested. The POSA would have been motivated to apply this teaching to use a shorter identifier site for Iafrate's method. Iafrate explained that for certain applications of its method the sample sizes would be small, including for the detection of mutations associated with particular cancer genes. EX1011 (Iafrate) at 0029. Comparing a small sample size to Kivioja's sequencing of the entire *Drosophila* transcriptome, the POSA would have been motivated to use fewer nucleotides in the molecular tag when applied to Iafrate's method, including 8 nucleotides, in order to extend the read length of the nucleic acid fragment. Indeed, the POSA would have been motivated to select the length of the molecular tag in order to optimize the number of available molecular tags—which would be 4^n, where "n" is the number of nucleotides in the molecular tag—relative to the expected number of nucleic acid fragments in a sample.

326. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 14.

327. I further note that the '399 patent does not teach that there was anything unexpectedly desirable or critical about an identifier site of specifically 8 nucleotides. To the contrary, the '399 patent does not provide any data whatsoever from experiments using the claimed methods. The POSA would not expect any such data—there is nothing particular special or valuable about an identifier site of 8

nucleotides (4^8, or 65,536 unique labels). That number of labels would be appropriate for some applications (with fewer unique fragments) but not appropriate for others (with more unique fragments), and the POSA would recognize that the determination of the appropriate length of the identifier site would be a matter of routine optimization, as I explain above in ¶¶ 264-266.

l. Claim 15: "The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site."

328. Claim 1 would have been obvious to the POSA for the reasons explained above.

329. Claim 15 would have been obvious based on Iafrate and Kivioja. The '399 patent defines a universal primer as "a primer that can be used for amplifying a number of different sequences." EX1003 ('399 patent) at 5:57–58.

330. Both Iafrate and Kivioja used standard Illumina sequences in their respective adaptors. EX1011 (Iafrate) at 0011; EX1012 (Kivioja) at 0014. The POSA would have known that the sequence in **blue** below (referred to as the "first sequencing platform primer" that "contains [a] barcode sequencing primer binding site"), EX1011 (Iafrate) at 0011, was the same in all of Iafrate's adaptors. Thus, that sequence was "used for amplifying a number of different sequences" and was a universal indexing primer binding site as that term is defined in the '399 patent.

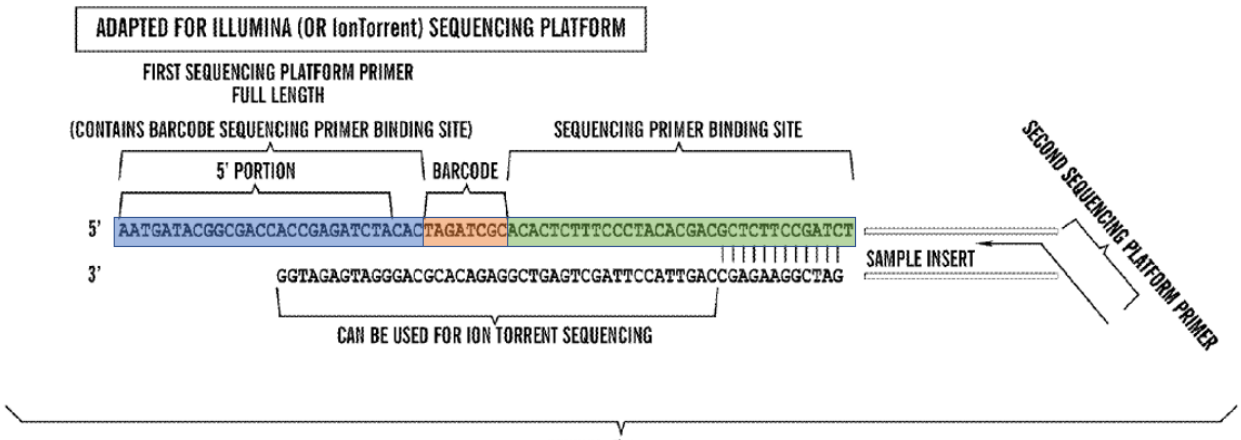


FIG. 9

331. The POSA also would have known that the **green** sequence in the figure below was a universal Illumina sequencing primer binding site that, in Kivioja's method, also functioned as the sample index primer binding site. That sequence was likewise the same in all of the adaptors and thus was "used for amplifying a number of different sequences," and was accordingly "universal" as that term is used in the '399 patent.

Supplementary Table 1. Sequences for oligomers used in mRNA-seq. Sequences in square brackets and parentheses represent an index sequence that was used to enable multiplexed sequencing, and a random ten base DNA sequence label, respectively. dU and rG represent deoxyuridine and guanine ribonucleotide, respectively.

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

332. Moreover, it would have been obvious to the POSA to use universal sample index primer binding sites in adaptors used for multiplex sequencing, as the

universal sequences would have been cost-effective and easily applied to NGS platforms, like Illumina. There would have been no reason to use variable primer binding sites, as this would have simply added complexity to the SBS process.

333. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 15.

m. Claim 16: "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site."

334. Claim 1 would have been obvious to the POSA for the reasons explained above.

335. Claim 16 would have been obvious based on Iafrate and Kivioja. . The '399 patent defines a universal primer as "a primer that can be used for amplifying a number of different sequences." EX1003 at 5:57–58.

336. As described and shown above in Section VI.C.6.l with respect to claim 15 above, both Iafrate and Kivioja included the same target sequence primer binding site (in **green**). EX1011 (Iafrate) at 0011; EX1012 (Kivioja) at 0014. The POSA would have recognized that this sequence was the same in all of the adaptors and was used for "amplifying a number of different sequences." Thus, this sequence was a "universal" primer binding site as that term is defined in the '399 patent.

337. Moreover, the POSA would have had no reason to use multiple different sequencing primer binding sites for a given sequencing read in a method

for NGS library preparation for Illumina. Using multiple primers for a given sequencing read for the SBS reaction would have been a departure from standard Illumina workflows and would have simply made the sequencing process unnecessarily more complicated.

338. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 16.

n. Claim 17: "The method of claim 6, wherein the method further comprises separating fragments or sequencing reads based on indexing site."

339. Claims 1, 5, and 6 would have been obvious to the POSA for the reasons I discuss above.

340. Claim 17 would have been obvious based on Iafrate and Kivioja. As I described above in Sections VI.C.6.a.2 and VI.C.6.a.5, Iafrate and Kivioja both describe the use of sample indexes for multiplex sequencing. EX1011 (Iafrate) at 0011, 0013, 0029, 0052; EX1012 (Kivioja) at 0003, 0014. Pooling samples for a single sequencing run would have required separating fragments or sequencing reads based on the sample index.

341. The POSA also would have known using sample indexes for multiplexed sequencing, followed by separation of samples (fragments or sequencing reads) based on indexing site, was consistent with the standard Illumina process in 2013. The process of separating based on indexing site would have been

referred to as "demultiplexing." EX1060 (Meyer & Kircher) at 0009; EX1061 (Kircher Thesis) at 0053; EX1062 (Smith) at 0034. After a multiplexed sequencing run, Illumina devices would have output sequencing data in FASTQ files that separated the sequenced nucleic acid fragments by their associated sample indexes. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034-35.

342. For these reasons, the POSA further would have had a reasonable expectation of success in practicing the method of claim 17.

7. Claims 1-4, 6-7, And 9-10 Of The '357 Patent Would Have Been Obvious Over Iafrate And Kivioja.

343. As I explain further below, the reasons why claim 1 of the '399 patent would have been obvious over Iafrate and Kivioja apply equally to claim 1 of the '357 patent. Each limitation recited in claim 1 of the '357 patent is also in claim 1 of the '399 patent, with the '357 patent being broader:

’357 Claim 1	’399 Claim 1
The invention claimed is: 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.	1. 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.

344. Several of the dependent claims of the ’357 patent are near copies of certain dependent claims in the ’399 patent, and therefore, as further explained below, also would have been obvious over Iafrate and Kivioja:

’357 Patent Dependent Claims	’399 Patent Dependent Claims
2. The method of claim 1, further comprising removing a duplicate read from the sequence reads. EX1004 at 25:48–49.	2. 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. EX1003 at 25:62–65.
6. The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1004 at 26:39–40.	10. The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1003 at 26:53–54.

'357 Patent Dependent Claims	'399 Patent Dependent Claims
7. The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site. EX1004 at 26:41–42.	16. The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site. EX1003 at 26:65–67.
9. The method of claim 1, wherein each unique identifier is 8 nucleotides in length. EX1004 at 26:45–46.	14. The method of claim 1, wherein the identifier site is 8 nucleotides in length. EX1003 at 26:61–62.

345. The remaining dependent claims of the '357 patent that do not have direct overlap with the '399 patent—claims 3, 4, and 10—are also obvious over Iafrate and Kivioja, as I explain further below.

a. Claim 1

346. Claim 1 is reproduced below:

> The invention claimed is:
> **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.

347. For the reasons I explain in relation to the '399 patent in Section VI.C.6.a of my declaration, claim 1 of the '357 patent would have been obvious to the POSA at the priority date based on Iafrate and Kivioja. The POSA would have been motivated to use Iafrate's method of library preparation where target

enrichment prior to sequencing was desired, e.g. for cancer research. The POSA further would have been motivated to improve Iafrate's method, which would have confounded accurate counting of original fragments and been hampered by PCR and sequencing error, by incorporating Kivioja's molecular tags and use of UMIs. Based on this combination, the POSA would have been motivated to practice a method comprising all of the recited steps of claim 1 of the '357 patent and use an adaptor with all of the claimed features.

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

348. The preamble to claim 1 in the '399 patent and the preamble to claim 1 in the '357 patent recite identical limitations. The POSA would have understood that there was no substantive difference between the two preambles, even though the '357 preamble is plural ("detecting duplicate sequencing reads") and the '399 preamble is singular ("detecting a duplicate sequencing read").

'357 Preamble	**'399 Preamble**
1. "A method for detecting duplicate sequencing reads" EX1004 at 25:36.	1. "A method for detecting a duplicate sequencing read from a population of sample sequencing reads" EX1003 at 25:37–38.

349. Because the '357 patent is directed to a method that is also claimed in the '399 patent, my opinion in Section VI.C.6.a.1 applies equally here. Specifically, the POSA would have recognized that Kivioja teaches a method of detecting and

removing duplicate sequencing reads.

2) "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"

350. The first step in claim 1 of the '357 patent overlaps with the first step in claim 1 of the '399 patent, as shown in the table below.

'357 Ligation Step	'399 Ligation Step
"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" EX1004 at 25:38–42.	"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 each 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" EX1003 at 25:39–46.

351. Each limitation in the ligation step in the '357 patent, including the features of the adaptor, is also present in the '399 patent. The differences between the two are that the '357 limitations are broader. First, while the '399 patent requires ligation "to a 5′ end of each nucleic acid fragment," EX1003 at 25:39, the '357 patent simply requires ligation to nucleic acid fragments, whether it be on the 5′ or 3′ end, EX1004 at 38-39. Second, while the '399 patent requires that adaptors feature specific primer binding sites, "an indexing primer binding site" and "a target

sequence primer binding site," EX1003 at 25:42, 46, the '357 patent requires simply that adaptors feature "a primer binding site," EX1004 at 25:42. Third, while the '399 patent requires the identifier site to be "between 1 and 8 nucleotides," EX1003 at 25:44, the '357 patent requires the identifier site to be "about 1 to about 8 nucleotides," EX1004 at 25:40.

352. I do not understand the '357 patent's characterization of an identifier site as "a unique identifier," EX1004 at 25:39, to be different than the "identifier site" in the '399 patent, EX1003 at 25:44. The '357 patent does not define "unique identifier;" rather, it defines "identifier site," EX1004 at 4:49–52, in the same way the '399 patent does, EX1003 at 4:43–46. And certain dependent claims in the '357 patent use "identifier site" rather than "unique identifier." EX1004 at 25:50, 26:35.

353. Rather, I understand the qualifier "unique" to mean that "the sequence of the identifier site is variable in sequence content" in each adaptor, as is required in claim 1 of the '399 patent. EX1003 at 25:59–61.

354. In any event, the identifiers used in Kivioja were random and thus used as "unique" identifiers of particular fragments as I discuss in Section VI.C.4 above.

355. In my opinion, because each limitation in the ligation step of the '357 patent is also present in the '399 patent, the ligation step in the '357 patent, including the features of the adaptor, would have been obvious based on the combination of Iafrate and Kivioja, for the reasons I discuss in Sections VI.C.6.a.2 and VI.C.6.a.5-

6 of my declaration.

3) "amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences"

356. The second and third steps in claim 1 of the '357 patent overlap with the second and third steps in claim 1 of the '399 patent, as shown in the table below.

'357 Amplification and Sequencing Step	**'399 Amplification and Sequencing Step**
"amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences" EX1004 at 25:43–45.	"b) amplifying the adapter-nucleic acid fragment ligated products; c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products" EX1003 at 25:47–51.

357. Despite slight differences in the language, there is no substantive difference between these steps of the '357 and '399 patents. The '399 patent does not use the term "amplicons," but the POSA would have understood the ordinary meaning of "amplicons" to be the amplified adaptor-ligated nucleic acid fragments. (The '357 patent does not provide a specific definition for or otherwise define "amplicons.")

358. The POSA also would have understood "sequencing the amplicons to produce sequence reads," EX1004 at 25:44, to be equivalent to "generating a population of sequencing reads," EX1003 at 25:49, *i.e.* both the '357 and the '399 patents require sequencing after amplification.

359. Finally, while the ’399 patent does not expressly say that sequencing reads “include identifier and target sequences,” as the ’357 patent does, EX1004 at 25:46–47, the ’399 patent necessarily requires it. Otherwise remaining steps of the ’399 patent would not have worked; for example, without the identifier and target sequences, duplicate reads could not be identified as “comprising the same identifier site and nucleic acid fragment.” EX1003 at 25:52–55.

360. Because each limitation in the amplification and sequencing steps of the ’357 patent are also present in the ’399 patent, these limitations in the ’357 patent would have been obvious based on the combination of Iafrate and Kivioja, for the reasons I explain in Section VI.C.6.a.3 of my declaration.

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

361. The fourth step in claim 1 of the ’357 patent overlaps with the fourth step in claim 1 of the ’399 patent, as shown in the table below.

’357 Identification Step	’399 Identification Step
“identifying sequence reads with identical identifier and target sequences as duplicates” EX1004 at 25:46–47.	“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” EX1003 at 25:52–55.

362. Despite slight differences in the language, both the ’357 patent and the ’399 patent require the same thing: identify duplicate sequencing reads as having

identical (or the same) identifier site and nucleic acid fragment sequences. The POSA would have understood "identical," as used in the '357 patent, EX1004 at 25:46, to be equivalent to "same," as used in the '399 patent, EX1003 at 25:43. In other words, the POSA would have understood that "identical identifier and target sequences," EX1004 at 25:46–47, means the same thing as "the same identifier site and nucleic acid fragment," EX1003 at 25:53–54. The '357 patent even describes duplicate sequencing reads in the language of the '399 patent: "A duplicate sequencing read can be a sequencing read with the same identifier site and target sequence as another sequencing read in the population of sequencing reads." EX1004 at 4:12–15.

363. Because this limitation in the '357 patent is also present in the '399 patent, the identification step in the '357 patent would have been obvious based on the combination of Iafrate and Kivioja, for the reasons I explain in Section VI.C.6.a.4 of my declaration.

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

364. Claim 1 would have been obvious to the POSA for reasons explained above.

365. Claim 2 of the '357 patent claims the same limitation as claim 2 of the '399 patent, as shown in the table below.

’357 Claim 2	’399 Claim 2
2. “The method of claim 1, further comprising removing a duplicate read from the sequence reads.” EX1004 at 25:48–49.	2. “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.” EX1003 at 25:61–64.

366. Because the limitation in claim 2 of the ’357 patent claim is equivalent to the limitation in claim 2 of the ’399 patent, claim 2 of the ’357 patent would have been obvious for the reasons I explain in Section VI.C.6.b of my declaration.

c. Claim 3: “The method of claim 1, wherein the identifier site is sequenced with the target sequence.”

367. Claim 1 would have been obvious to the POSA for the reasons explained above.

368. As I explained in relation to claim 3 of the ’399 patent, the ordinary meaning to the POSA of sequencing the identifier site with the target sequence is that both sequences are sequenced in a single SBS read on the NGS device. Under that interpretation, claim 3 would have been obvious based on Iafrate and Kivioja. Kivioja constructed the 5′ adaptor with the molecular tag and the nucleic acid fragment on the 3′ end of the sequencing primer binding site. EX1012 (Kivioja) at 0014. And as I explained above in Sections VI.C.4–5, the POSA would have been motivated to modify Iafrate’s method to add the molecular tag taught by Kivioja to the 3′ side of the sequencing primer binding site in Iafrate’s adaptor. The placement

of the sequencing primer binding site relative to the molecular tag and nucleic acid fragment means the molecular tag and nucleic acid fragment would have been sequenced in a single SBS read.

369. For Kivioja's mRNA experiment, the adaptor-cDNA fragments were sequenced in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at 0004. The POSA would have understood Kivioja's disclosure of 54-bp reads to mean the identifier site was sequenced with the nucleic acid fragment because Kivioja discloses only a single read of the adaptor-cDNA fragment.

370. The POSA's understanding would have been confirmed by Kivioja's separate disclosures of paired-end sequencing in the genomic DNA experiment and other mRNA experiment. For the gDNA experiment, Kivioja explained that "sequences were generated on an Illumina Genome Analyzer with 76-bp single reads for the mixed sample and 100-bp paired-end reads for the adult male sample." EX1012 (Kivioja) at 0004. In the other mRNA experiment, Kivioja explained that the samples were sequenced on an Illumina HiSeq 2000 with "54-nucleotide reads from both ends." EX1012 (Kivioja) at 0004. With the molecular tag in the 5′ adaptor and 3′ to the sequencing primer binding site—i.e., an adaptor structure of 5′-Sequencing Primer Binding Site–Molecular Tag–Sample Index–Fragment—paired end reads would have resulted in the identifier site and fragment being sequenced together in one direction, but not the other. Kivioja, moreover, distinguished

between the use of single read sequencing and paired-end sequencing, as in the DNA experiment. EX1012 (Kivioja) at 0004. The POSA would thus have understood Kivioja to teach that the molecular tag and nucleic acid fragment in one mRNA experiment were sequenced together in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at 0004.

371. To the extent that claim 3 requires the identifier site to be placed in the adaptor in such a position as to allow the identifier site and the nucleic acid fragment to be sequenced together in a single SBS read, the choice of where to place the molecular tag in any particular adaptor would have been a design choice for the POSA. Indeed, as I explained above in Section VI.C.5 of my declaration, given the teachings of Iafrate and Kivioja, the POSA would have inserted the random molecular tag on the 3′ side of the sequencing primer binding site, as I show below in an annotated comparison of Figure 9 from EX1011 (Iafrate) and Supplementary Table 1 from EX1012 (Kivioj), below. EX1011 (Iafrate) at 0011; EX1012 (Kivioja) at 0014.

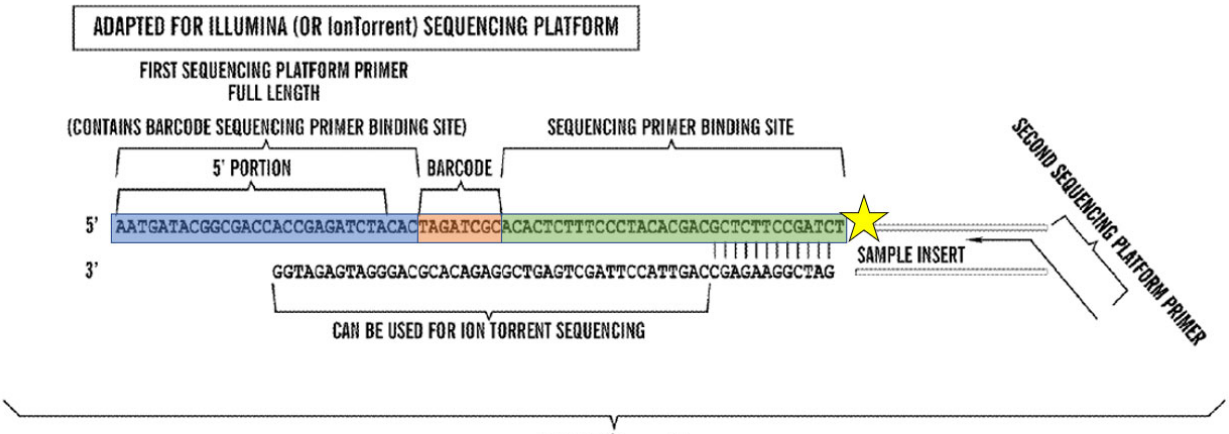


FIG. 9

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

372. Placing the molecular tag to the 3′ side of the sequencing primer binding site means that the molecular tag and nucleic acid sequence could have been sequenced in a single SBS read.

373. For certain Illumina devices, the molecular tag could be read from the sequence primer binding site. Thus, the POSA would have known that the molecular tag and sequencing primer binding site could be placed in different positions within an adaptor depending on the requirements of an experiment such that the molecular tag and nucleic acid fragment were sequenced in a single SBS read.

374. For these reasons, the POSA would have had a reasonable expectation of success in practicing the method of claim 3.

375. I further do not understand the '357 patent to disclose anything

unexpected or beneficial about the placement of the identifier site within the adaptor so as to sequence the identifier site and target sequence in one SBS read or two. To the contrary, the POSA would have recognized that useful sequencing data could be obtained whether the identifier site was sequence with or without the target sequence.

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence."

376. Claim 1 would have been obvious to the POSA for the reasons explained above.

377. The ordinary meaning to the POSA of sequencing the identifier site "separately" from the target sequence in this context is that each sequence is sequenced in a separate SBS read on the NGS device. As I explained above in Section VI.C.7.d, Kivioja disclosed the use of paired-end sequencing in the genomic DNA experiment and one mRNA experiment. For the DNA experiment, Kivioja explained that "sequences were generated on an Illumina Genome Analyzer with 76-bp single reads for the mixed sample and 100-bp paired-end reads for the adult male sample." EX1012 (Kivioja) at 0004. In the other mRNA experiment, Kivioja explained that the samples were sequenced on an Illumina HiSeq 2000 with "54-nucleotide reads from both ends." EX1012 (Kivioja) at 0004. With the molecular tag only in the 5′ adaptor, paired-end sequencing would have generated one read

from the 5′ side that contained the molecular tag sequence and another read from the 3′ side that sequenced the nucleic acid fragment separately without the molecular tag, at least for fragments longer than the maximum read length of 54 nucleotides. The POSA would have understood Kivioja to disclose a method wherein the molecular tag and nucleic acid fragment were sequenced separately.

378. Furthermore, for reasons I explained above in Section VI.C.7.c, to the extent claim 4 requires the identifier site to be placed in a position within an adaptor such that the identifier site and nucleic acid fragment were sequenced separately, that would have been a design choice for the POSA. It would have been routine at the priority date to optimize the structure of the adaptor and the modes of the Illumina device for a particular application. The POSA would have been motivated to sequence the label sequence separately from the fragment sequence for some applications, e.g. in order to have a short first read to sequence the label followed by a long second read through the fragment to maximize the length of fragment sequenced. Moreover, the POSA would have had a reasonable expectation of success in sequencing the label and the fragment separately, as this was standard in paired-end sequencing at the priority date. In addition, I do not understand the ’357 patent to disclose anything unexpected or important about sequencing the label and the fragment separately instead of in a single read—to the contrary, the POSA would have recognized that these were both routine choices depending on the experimental

design.

379. For these reasons, the POSA would have had a reasonable expectation of success in practicing the method of claim 4.

e. Claim 6: "The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."

380. Claim 1 would have been obvious to the POSA for the reasons explained above.

381. Claim 6 of the '357 patent claims the same limitation as claim 10 of the '399 patent, as shown in the table below.

'357 Claim 6	'399 Claim 10
6. The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1004 at 26:39-40.	10. The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1003 at 26:53-54.

382. In the '399 patent, claim 10 depends from claim 9, which limits the nucleic acid fragments to "DNA fragments, RNA fragments, or DNA/RNA fragments." EX1003 at 26:33–34. So claim 10 of the '399 patent limits claim 9 to more specific types of DNA fragments, *i.e.* genomic DNA and cDNA fragments. Claim 6 of the '357 patent provides for the same specificity but it depends from claim 1. I understand that these differences do not alter the substance of what is claimed, and thus I understand claim 6 of the '357 patent to claim the same limitation as claim 10 of the '399 patent.

383. Because the limitation in claim 6 of the ’357 patent is equivalent to the limitation in claim 10 of the ’399 patent, claim 6 of the ’357 patent would have been obvious for the reasons I gave above in Section VI.C.6.i of my declaration.

f. Claim 7: “The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site.”

384. Claim 1 would have been obvious to the POSA for the reasons explained above.

385. Claim 7 of the ’357 patent overlaps with the limitation in claim 16 of the ’399 patent, as shown in the table below.

’357 Claim 7	’399 Claim 16
7. “The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site.” EX1004 at 26:41–42.	16. “The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.” EX1003 at 26:65–67.

386. Both claim 7 of the ’357 patent and claim 16 of the ’399 patent require “a universal target sequence primer binding site.” EX1004 at 26:41–42; EX1003 at 26:65–67. That claim 7 of the ’357 patent limits a “primer binding site,” as opposed to “the target sequence primer binding site,” as in the ’399 patent, does not alter my opinion because both ultimately specify that a primer binding site in the adaptor must be “a universal target sequence primer binding site.”

387. Thus, because the limitation in claim 7 of the ’357 patent is equivalent to the limitation in claim 16 of the ’399 patent, claim 7 of the ’357 patent would have

been obvious for the reasons I provided in Section VI.C.6.m of my declaration.

g. Claim 9: "The method of claim 1, wherein each unique identifier is 8 nucleotides in length."

388. Claim 1 would have been obvious to the POSA for the reasons explained above.

389. Claim 9 of the '357 patent claims the same limitation as claim 14 of the '399 patent, as shown in the table below.

'357 Claim 9	'399 Claim 14
9. The method of claim 1, wherein each unique identifier is 8 nucleotides in length. EX1004 at 26:45–46.	14. The method of claim 1, wherein the identifier site is 8 nucleotides in length. EX1003 at 26:61–62.

390. Because the limitation in claim 9 of the '357 patent is equivalent to the limitation in claim 14 of the '399 patent, claim 9 of the '357 patent would have been obvious for the reasons I provided in Section VI.C.6.k of my declaration.

h. Claim 10: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic acid fragment."

391. Claim 1 would have been obvious to the POSA for the reasons explained above.

392. The '357 patent does not define "junction" but the ordinary meaning to the POSA in this context is the point where the two sequences are joined. The '357 patent uses the term "junction" in this way; for example, in Figure 4 where the identifier site in the adaptor is joined to the nucleic acid fragment at what is described

as "the insert junction." EX1004 ('357 patent) at 0010. Thus, I understand claim 10 of the '357 patent to require the identifier site to be located in the adaptor such that the identifier site is joined to the nucleic acid fragment.

393. For reasons I gave in Sections VI.C.5 and VI.C.7.c, in my opinion, based on the combination of Iafrate and Kivioja, it would have been obvious to the POSA to put the molecular tag on the 3′ side of the adaptor in Iafrate's Figure 9, as shown in the annotated figure below. EX1011 (Iafrate) at 0011. Placing the molecular tag in the starred position below would have put the molecular tag "at a junction between the adapter and the nucleic acid fragment." EX1004 at 26:47–49. This is shown below in my annotated comparison of Figure 9 from EX1011 (Iafrate) and Supplementary Table 1 from EX1012 (Kivioja). EX1011 (Iafrate) at 0011; EX1012 (Kivioja) at 0014.

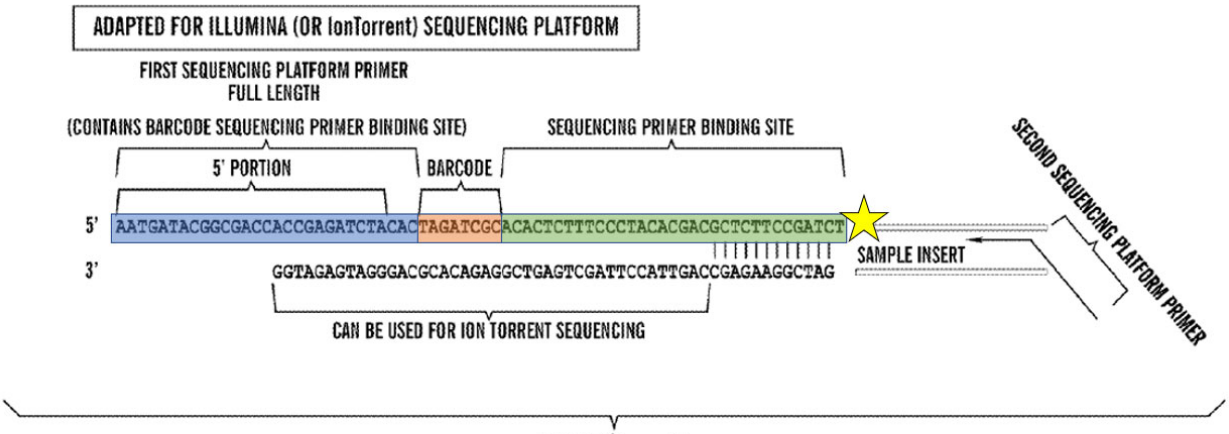


FIG. 9

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

8. Claims 1-16 Of The '241 Patent Were Anticipated By Kivioja.

394. Each limitation recited in claim 1 of the '241 is also present in claim 1 of the '399 patent, and it is even broader than the '357 patent. There is also substantial overlap between the '241 patent's dependent claims and the claims of the '399 and '357 patents. Because of the breadth of the claims of the '241 patent, claims 1-16 were anticipated by Kivioja.

a. Claim 1

395. Claim 1 is reproduced below:

> The invention claimed is:
> **1**. A method for detecting duplicate sequencing reads, the method comprising:
> obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor, wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment;
> sequencing the amplicons to generate sequence reads that include identifier and target sequences; and
> identifying sequence reads with identical identifier and target sequences as duplicates.

396. In my opinion, claim 1 was anticipated by Kivioja. As I explain in this section, Kivioja discloses every step of the method, including as shown in Figure 1 and the associated text, and also in Kivioja's mRNA proof of concept experiment.

1) "A method for detecting a duplicate sequence read from a population of sample sequencing reads"

397. As shown in the table below, the preamble to claim 1 in the '241 patent and the preamble to claim 1 in the '399 and '357 patents are directed to the same method. The POSA would have understood that there was no substantive difference between the preambles, even though the '241 and '357 preambles are plural (i.e., "detecting duplicate sequencing reads") and the '399 preamble is singular (i.e., "detecting a duplicate sequencing read"). The POSA would understand that the use of an NGS device and the associated data processing in Kivioja necessarily applied to a group of sequencing reads, and thus also to individual sequencing reads within the population.

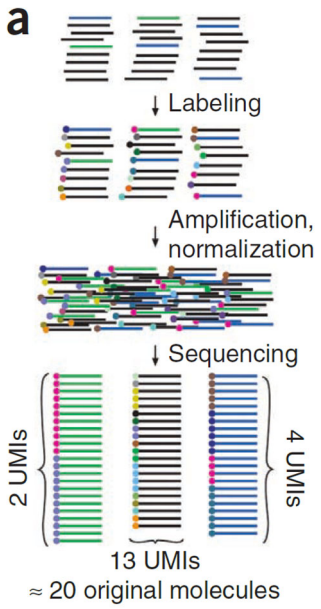
’241 Preamble	’357 Preamble	’399 Preamble
1. “A method for detecting duplicate sequencing reads” EX1005 at 25:33–34.	1. “A method for detecting duplicate sequencing reads” EX1004 at 25:36.	1. “A method for detecting a duplicate sequencing read from a population of sample sequencing reads” EX1003 at 25:37–38.

398. Because the ’241 claim 1 preamble is substantively the same as the ’357 and ’399 preambles that I discuss above, my explanation in Section VI.C.6.a.1 applies equally here. Specifically, the POSA would have recognized that Kivioja teaches a method of detecting and removing duplicate sequencing reads.

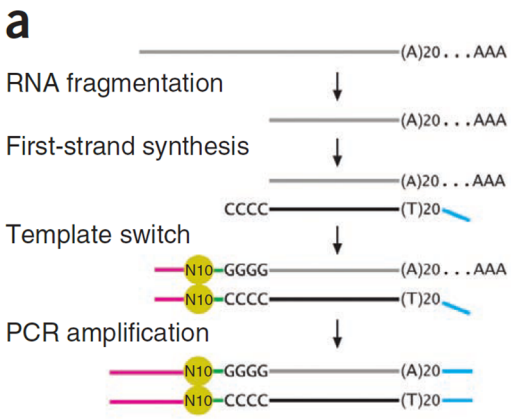
2) “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor”

399. Claim 1 next recites “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor.” EX1005 (’241 patent) at 25:36–37. For the reasons I discuss below, Kivioja discloses this limitation.

400. Kivioja teaches appending adaptors with random tags to fragments prior to sequencing and then amplifying those labeled fragments, as shown in Figure 1a below. EX1012 (Kivioja) at 0001, Figure 1a. In the third step of EX1012 (Kivioja) Figure 1a, which I have reproduced below, Kivioja obtains a set of amplicons that each comprise an amplified fragment of nucleic acid with an appended adaptor.



401. Kivioja exemplified this concept in the *Drosophila* mRNA example. Specifically, after mRNA fragmentation, Kivioja added an adaptor to the 5' end of each fragment using a template switch reaction. EX1012 (Kivioja) at 0002. Kivioja depicts this in Figure 3a, reproduced below. EX1012 (Kivioja) at 0002. The added adaptor includes an Illumina linker sequence in magenta, followed by a random 10 base sequence label in gold ("N10"), and a sample index in green. EX1012 (Kivioja) at 0002.



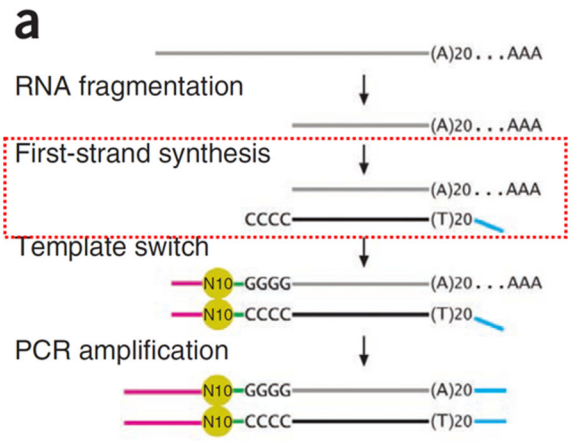
402. Kivioja discloses the sequence of this adaptor in Supplementary Table 1, reproduced below. EX1012 (Kivioja) at 0014. As Kivioja describes, the Ns in the adaptor sequence are "a random ten base DNA sequence label." EX1012 (Kivioja) at 0014.

5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG

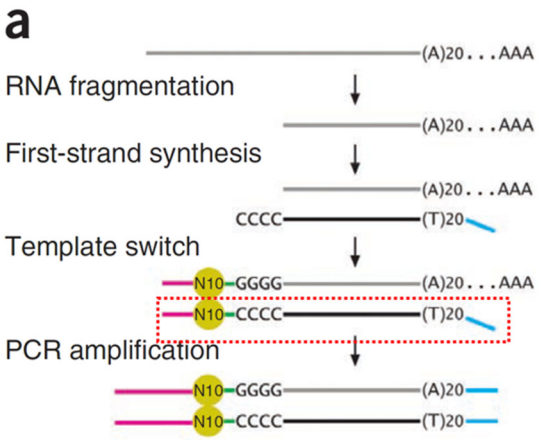
403. Template switching is a method of appending an adaptor. In my opinion, the term "appended" does not place any limitation on the method that may be used to attach the adaptor to the fragment, and the '241 patent does not define the term. Thus, the POSA would understand that the term would not place any limitation on the process by which the adaptor was attached and the adaptor could be attached by any means.

404. Moreover, the '241 patent explains that "[i]n some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction." EX1005 ('241 patent) at 0016 at 9:34–36. Because this disclosure says it describes "some embodiments" the POSA would understand that it does not exclude other means of attaching the adaptor. But regardless, the template switch method used by Kivioja is a type of priming reaction, which is one of the two types of methods, in addition to ligation, of "appending" specifically disclosed in the '241 patent. EX1005 ('241 patent) at 9:34–36.

405. In Kivioja, the starting samples were RNA, and the RNA was reverse transcribed prior to sequencing. EX1012 (Kivioja) 0004. Kivioja used a reverse transcriptase enzyme to generate a complementary strand of cDNA from the starting RNA template molecules (the "first strand synthesis" annotated in Figure 3a below), starting from a short primer with a noncomplementary adaptor tail (shown in blue in Figure 3a below) that annealed to one end of the RNA molecules. EX1012 (Kivioja) at 0002, 0004.

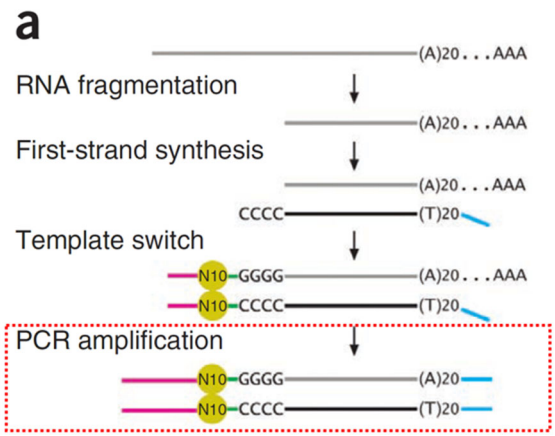


406. During the template switch method, Kivioja added the 5' adaptors to the reaction mixture. EX1012 (Kivioja) at 0002. When the reverse transcriptase reached the end of the RNA strand, the 5' adaptors anneal at the end of the complementary stands (specifically in the CCCC region), and the reverse transcriptase enzyme "switched" to use the 5' adaptor as the template. This switch generates a strand of cDNA complementary to the original RNA strand and the 5' adaptor, shown in the annotated version of Figure 3a below.



407. Because a primer is used to start the first strand synthesis reaction, and because the first strand of cDNA (modified with CCCC) itself serves as a primer for copying the 5' adaptor used in the template switching reaction, the POSA would have recognized that the mRNA template switch in Kivioja is a type of "priming reaction."

408. After the template switch and generation of the cDNA strand, the Kivioja method amplifies the fragments via PCR to obtain a set of double stranded DNA molecules, or amplicons, shown in the annotated Figure 3a below. The amplicon is comprised of the original fragment sequence with the appended adaptors, including the random tag.



409. At the end of this process, Kivioja obtained a library of amplified fragments (the gray and black bars in the figure above) with appended adaptors (the colored bars on either end of the fragments in the figure above) that contained the identifier site ("N10").

3) "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment"

410. Claim 1 next recites "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment. EX1005 ('241 patent) at 25:37–39. Kivioja discloses this limitation.

411. Kivioja explains that its method involves the attachment of "random labels" to each fragment that "can be used to generate a library in which each molecule has a distinct identifying sequence." EX1012 (Kivioja) at 0001. The random labels are an "identifier site" that is comprised of a plurality of nucleotides that are "distinct" or "unique" to that particular fragment.

412. The Kivioja mRNA template method uses a "10-base pair random label" that uniquely labels each RNA fragment. EX1012 (Kivioja) at 0002, 0014. The Kivioja method confirmed after sequencing that there are sufficient random labels to ensure that each fragment has a label "unique to the amplified fragment." Kivioja specifically explains that "[t]he approximately one million random labels were sufficient to generate UMIs [*i.e.*, unique adaptor-fragment concatenations]" from the mRNA samples. EX1012 (Kivioja) at 0005. Kivioja further explains that "the UMIs were not exhausted in the experiments." EX1012 (Kivioja) at 0005. In other words, there was an excess of unique tag sequences relative to fragment sequences, and thus the identifiers were unique to nearly every amplified fragment.

4) "sequencing the amplicons to generate sequence reads that include identifier and target sequence, and"

413. Claim 1 next recites "sequencing the amplicons to generate sequence reads that include identifier and target sequences, and" EX1005 ('241 patent) at 25: 40–41. Kivioja discloses this limitation for the reasons I discuss above in Sections VI.C.6.a.3 and VI.C.7a.3 of my declaration, as well as the reasons explained in this section.

414. This step in claim 1 of the '241 patent overlaps with the sequencing step in claim 1 of the '357 and '399 patents, as shown in the table below.

’241 Sequencing Step	’357 Sequencing Step	’399 Sequencing Step
“sequencing the amplicons to generate sequence reads that include identifier and target sequences” EX1005 at 25:40–41.	“sequencing the amplicons to produce sequence reads that include identifier and target sequences” EX1004 at 25:44–45.	“c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products” EX1003 at 25:44–51.

415. Despite slight differences in the language, there is no substantive difference between the steps of the ’241, ’357, and ’399 patents. The ’241 patent uses the term “generate,” and the ’357 patent uses the term “produce.” In this context, In my opinion, there is no substantive difference between the terms, and the POSA would have understood “generate sequence reads” to mean the same as “produce sequence reads.”

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

416. Claim 1 next recites “identifying sequence reads with identical identifier and target sequences as duplicates.” EX1005 (’241 patent) at 25:42–43. Kivioja discloses this limitations for the reasons I discuss above in Sections VI.C.6.a.4 and VI.C.7.a.4 of my declaration, as well as the reasons explained in this section.

417. This step in claim 1 of the ’241 patent is the same as the fourth step in claim 1 of the ’357 patent and substantively the same as the fourth step in claim 1 of the ’399 patent, as shown in the table below.

’241 Identification Step	’357 Identification Step	’399 Identification Step
“identifying sequence reads with identical identifier and target sequences as duplicates” EX1004 at 25:46–47.	“identifying sequence reads with identical identifier and target sequences as duplicates” EX1004 at 25:46–47.	“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” EX1003 at 25:52–55.

418. For the reasons I discuss above in Section VI.C.7.a.4, in this context, the POSA would have understood that there was no substantive difference between the claim limitation in claim 1 of the ’241 patent, claim 1 of the ’399 patent, and claim 1 of the ’357 patent. The POSA also would have understood, for the reasons I describe above in Sections VI.C.4, VI.C.6.a.4, and VI.C.7.a.4, that Kivioja teaches “identifying sequence reads with identical identifier and target sequences as duplicates.”

419. The POSA would have understood that there was no substantive differences between the claim limitation in claim 1 of the ’241 patent and claim 1 of the ’357 patent. The POSA also would have understood, for the reasons I describe above in Sections VI.C.6.a.4 and VI.C.7.a.4, that Kivioja teaches “identifying sequence reads with identical identifier and target sequences as duplicates.”

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

420. Claim 2 is anticipated by Kivioja. Claim 2 of the '241 patent recites the same limitation as claim 2 of the '399 and '357 patents, as shown in the table below.

'241 Claim 2	'357 Claim 2	'399 Claim 2
2. "The method of claim 1, further comprising removing a duplicate read from the sequence reads." EX1005 at [cite]	2. "The method of claim 1, further comprising removing a duplicate read from the sequence reads." EX1004 at 25:48–49.	2. "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." EX1003 at 25:61–64.

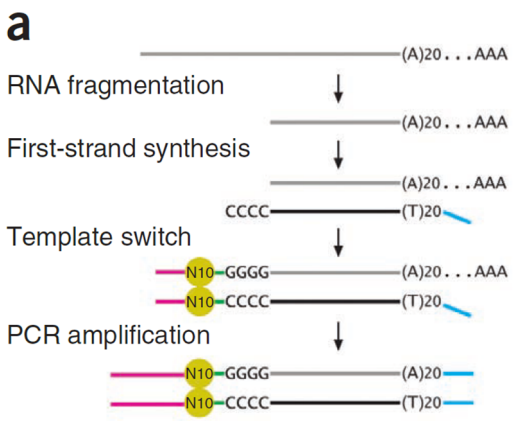
421. Because the limitation of claim 2 of the '241 patent claim is equivalent to the limitation of claim 2 of the '399 patent and claim 2 of the '357 patent, Kivioja discloses this limitation for the reasons stated in Sections VI.C.6.b and VI.C.7.b of my declaration.

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence."

422. Claim 3 is anticipated by Kivioja. Claim 3 of the '241 patent recites the same limitations as claim 3 of the '357 patent, as shown in the table below.

’241 Claim 3	’357 Claim 3
3. “The method of claim 1, wherein the identifier site is sequenced with the target sequence.” EX1005 at 25: 46–48.	3. “The method of claim 1, wherein the identifier site is sequenced with the target sequence.” EX1004 at 25:50–51.

423. As I explained in relation to claim 3 of the ’399 patent in Section VI.C.7.c, the ordinary meaning to the POSA of sequencing the identifier site with the target sequence is that both sequences are sequenced in a single SBS read on the NGS device. Under that interpretation, claim 3 is anticipated by Kivioja. Kivioja constructed the 5′ adaptor with the molecular tag and the nucleic acid fragment on the 3′ end of the sequencing primer binding site, as shown in Figure 3a below. EX1012 (Kivioja) at 0014. The placement of the sequencing primer binding site relative to the molecular tag and nucleic acid fragment means the molecular tag and nucleic acid fragment would have been sequenced in a single SBS read.



424. For Kivioja’s mRNA experiment, the adaptor-cDNA fragments were sequenced in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at

0004. The POSA would have understood Kivioja's disclosure of 54-bp reads to mean the identifier site was sequenced with nucleic acid fragment because Kivioja discloses only a single read of the adaptor-cDNA fragment.

425. The POSA's understanding would have been confirmed by Kivioja's separate disclosures of paired-end sequencing in the genomic DNA experiment and other mRNA experiment. For the gDNA experiment, Kivioja explained that "sequences were generated on an Illumina Genome Analyzer with 76-bp single reads for the mixed sample and 100-bp paired-end reads for the adult male sample." EX1012 (Kivioja) at 0004. In the other mRNA experiment, Kivioja explained that the samples were sequenced on an Illumina HiSeq 2000 with "54-nucleotide reads from both ends." EX1012 (Kivioja) at 0004. With the molecular tag in the 5′ adaptor and 3′ to the sequencing primer binding site—i.e., an adaptor structure of 5′-Sequencing Primer Binding Site–Molecular Tag–Sample Index–Fragment—paired end reads would have resulted in the identifier site and fragment being sequenced together in one direction, but not the other. Kivioja, moreover, distinguished between the use of single read sequencing and paired-end sequencing, as in the DNA experiment. EX1012 (Kivioja) at 0004. The POSA would thus have understood Kivioja to teach that the molecular tag and nucleic acid fragment in one mRNA experiment were sequenced together in 54-bp reads on an Illumina Genome Analyzer II. EX1012 (Kivioja) at 0004. As shown in the annotated figure below,

reproduced from EX1012.0014, the random molecular tag (in **yellow**) in Kivioja's adaptor was 3′ to the sequencing primer binding site (**green**).

Supplementary Table 1. Sequences for oligomers used in mRNA-seq. Sequences in square brackets and parentheses represent an index sequence that was used to enable multiplexed sequencing, and a random ten base DNA sequence label, respectively. dU and rG represent deoxyuridine and guanine ribonucleotide, respectively.

Oligomer	Sequence (5'-3')
Oligo-dT adapter	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG
Paired-end compatible oligo-dT adapter	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence."

426. Claim 1 is anticipated and/or rendered obvious by Kivioja. Claim 4 further requires that the identifier site and target sequence be sequenced separately. For the reasons I explain above in Section VI.C.7.d in connection with the identical claim 4 of the '357 patent, this limitation is disclosed in Kivioja and thus Kivioja anticipates claim 4 of the '241 patent, which depends from claim 1, which is anticipated for the reasons I explain above in Section VI.C.8.a.

427. Kivioja disclosed the use of paired-end sequencing in the genomic DNA experiment and one mRNA experiment. With the molecular tag only in the 5′ adaptor, paired-end sequencing would have generated one read from the 5′ side that contained the molecular tag sequence and another read from the 3′ side that sequenced the nucleic acid fragment separately without the molecular tag. The

POSA would have understood Kivioja to disclose a method wherein the molecular tag and nucleic acid fragment were sequenced separately.

428. At a minimum, as I explain above in Section VI.C.7.d, Kivioja would have made claim 4 obvious to the POSA. It would have been routine at the priority date to optimize the structure of the adaptor and the modes of the Illumina device for a particular application. The POSA would have been motivated to sequence the label sequence separately from the fragment sequence for some applications, *e.g.* in order to have a short first read to sequence the label followed by a long second read through the fragment to maximize the length of fragment sequenced. Moreover, the POSA would have had a reasonable expectation of success in sequencing the label and the fragment separately, as this was standard in paired-end sequencing at the priority date. In addition, I do not understand the '241 patent to disclose anything unexpected or important about sequencing the label and the fragment separately instead of in a single read—to the contrary, the POSA would have recognized that these were both routine choices depending on the experimental design.

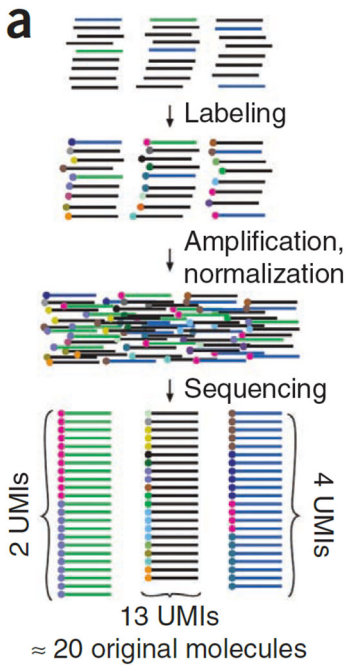
e. Claim 5: "The method of claim 1, wherein the amplicons include amplified fragments of cDNA."

429. Claim 5 is anticipated by Kivioja. Kivioja explains in its RNA example that the libraries for sequencing "were generated by fragmenting total RNA and reverse transcribing it to cDNA." EX1012 (Kivioja) at 0002.

f. Claim 6: "The method of claim 1, wherein the amplicons include amplified fragments of genomic DNA."

430. Claim 6 is anticipated by and/or obvious over Kivioja.

431. Kivioja anticipates claim 6 because it expressly teaches the use of genomic DNA in its methods. Figure 1, reproduced below, explains that "DNA species (green, blue, and black lines) are labeled with a collection of random labels (colored filled circles)." EX1012 (Kivioja) at 0001. These DNA samples are used in the subsequent steps of the Kivioja method, as I discuss above in Section VI.C.8.a. Kivioja also explains that its methods are applicable to "genomic DNA" specifically. EX1012 (Kivioja) at 0002.

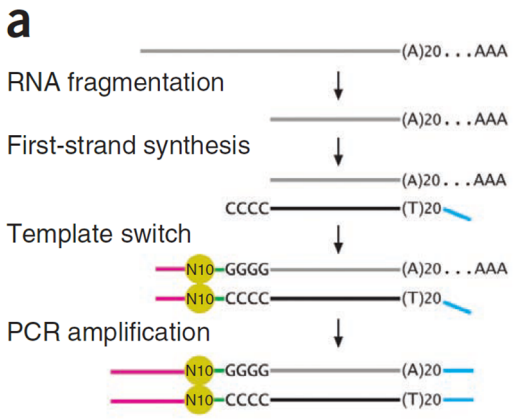


432. To the extent that Kivioja does not anticipate claim 6, at a minimum, Kivioja renders claim 6 obvious. The POSA would have recognized based on the

teachings of Figure 1 and Kivioja's examples that it would have been obvious to use Kivioja's methods with genomic DNA. EX1012 (Kivioja at 0001–0005). In addition, the POSA would have had a reasonable expectation of doing so, given Kivioja's express teachings that the methods were suitable for genomic DNA.

g. Claim 7: "The method of claim 1, wherein each adaptor includes a primer binding site."

433. Claim 7 is anticipated by Kivioja. Figure 3a of Kivioja, reproduced below, shows the preparation of the library of adaptor-ligated fragments. EX1012 (Kivioja) at 0002. The magenta sequence added in the adaptor was an "Illumina linker." EX1012 (Kivioja) at 0002. The POSA would have recognized that this linker contained a primer binding site.



434. Specifically, the Illumina linker served as a primer binding site for multiple primers, including a primer binding site in the PCR amplification reaction to generate the library of amplicons. EX1012 (Kivioja) at 0004. The magenta sequence in Figure 3a also included the primer binding site for the forward read

primer in the SBS reaction on the Illumina device, as I discuss further above in paragraph 242 above. EX1012 (Kivioja) 0014.

h. Claim 8: "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site."

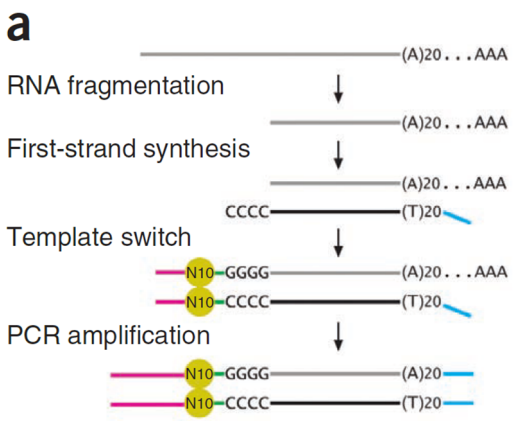
435. Claim 8 is anticipated by Kivioja. Claim 8 of the '241 patent recites the same limitations as claim 7 of the '357 patent and claim 16 of the '399 patent, as shown in the table below.

'241 Claim 8	'357 Claim 7	'399 Claim 16
8. "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site." EX1005 at 26:56–57.	7. "The method of claim 1, wherein the primer binding site is a universal target sequencing primer binding site." EX1004 at 42–43.	16. "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site." EX 1003 at 65–67.

436. Thus, Claim 8 of the '241 patent is anticipated for the reasons I discuss above in connection with claim 16 of the '399 patent in Section VI.C.6.m above, and claim 7 of the '357 patent in Section VI.C.7.f above.

i. Claim 9: "The method of claim 1, wherein teach unique identifier is at least 6 nucleotides in length."

437. Claim 9 is anticipated by Kivioja. Kivioja used an identifier of at least 6 nucleotides in its mRNA example, as shown in Figure 3a, reproduced below. The gold "N10" is a "10-base-pair random label." EX1012 (Kivioja) at 0002.



438. This is consistent with the sequence of the 5′ adaptor in Supplementary Table 1, reproduced below. EX1012 (Kivioja) at 0014. As Kivioja describes, the Ns in the adaptor sequence are "a random ten base DNA sequence label." EX1012 (Kivioja) at 0014.

5'-adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(NNN)dU(NNNN)dU(NNN)[GACTT]rGrGrGrG

j. Claim 10: "The method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction."

439. Claim 10 is anticipated by Kivioja. For the reasons I explain above in Section VI.C.8.a.2, Kivioja teaches appending adaptors to fragments using a template switch, a kind of priming reaction.

440. In addition, Kivioja discloses ligating adaptors to fragments prior to sequencing. EX1012 (Kivioja) at 0004. The POSA would have understood that the labels described in, for example, Figure 1a, of Kivioja could have been attached via ligation.

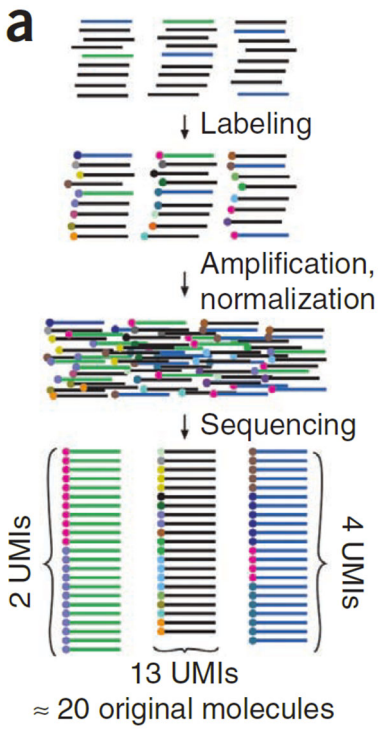
k. Claim 11: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment."

441. Claim 11 is anticipated by and/or obvious over Kivioja. Claim 11 of the '241 patent recites the same limitations as claim 10 of the '357 patent, as shown in the table below.

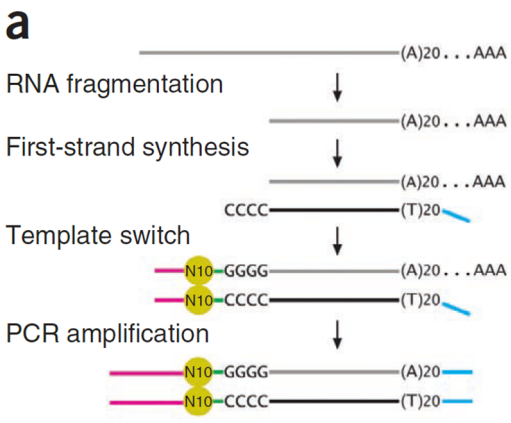
'241 Claim 11	'357 Claim 10
11. "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment." EX1005 at 26:35–37.	10. "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment." EX1004 at 26:48–50.

442. Kivioja anticipates this limitation and/or renders it obvious for the same reasons I discuss above in connection with '357 claim 10 in Section VI.C.7.h above.

443. Specifically, in Figure 1a as shown below, Kivioja shows the identifier/label sequences, the colored filled circles, at the junction with the fragments, the green, blue, and black lines. EX1012 (Kivioja) at 0001. Thus, Kivioja expressly discloses the limitation of claim 11 and thus anticipates it.



444. To the extent that claim 11 is not expressly disclosed, at a minimum Kivioja renders claim 11 obvious. In the mRNA example, Kivioja placed the sample index at the junction between the adapter and the nucleic acid fragment, as shown in Figure 3a.



445. But it would have been obvious to the POSA that the sample index and

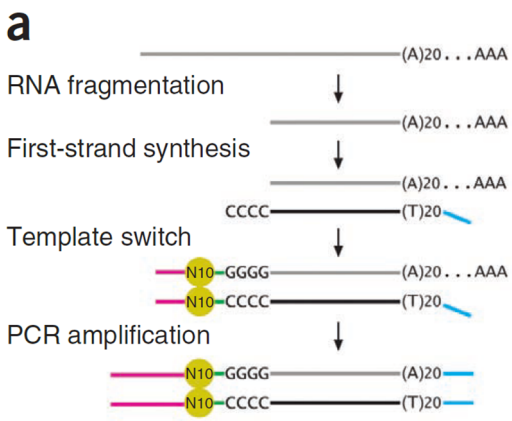
the molecular tag could be swapped in the order of the adaptor. Either of these two options would be equally suitable to the POSA's goals. Thus, it would have been obvious to the POSA to consider either position for the identifier site and the POSA would have had a reasonable expectation of success in swapping the order of the identifier and the sample index. In addition, I do not understand the '399 patent to disclose any unexpected properties of placing the identifier site at the junction with the fragment as opposed to at any other position in the adaptor.

l. Claim 12: "The method of claim 1, further comprising determining the genomic start position for each sequencing read."

446. Kivioja anticipates claim 12. Kivioja expressly teaches determining the genomic start position for each read. Specifically, Kivioja explains that "the mapped genomic position of either end can be used as an UMI." EX1012 (Kivioja) at 0002. In connection with the mRNA example specifically, Kivioja explains that "[t]he combination of the label sequence and the 5′ mapped position of the RNA fragment forms the UMI." EX1012 (Kivioja) at 0002, 0003. The 5′ mapped position is the genomic start position. This is consistent with Kivioja's detailed description of the data analysis, in which the authors wrote, "[t]he mapped reads with the same gene, position, and label were collected to one UMI." EX1012 (Kivioja) at 0005.

m. Claim 13: "The method of claim 1, wherein each adaptor further comprises an indexing site."

447. Kivioja anticipates claim 13. In its mRNA example, Kivioja states that the 5′ template switch adaptor included an "index sequence" shown in green in the excerpted Figure 3a below. EX1012 (Kivioja) at 0002.



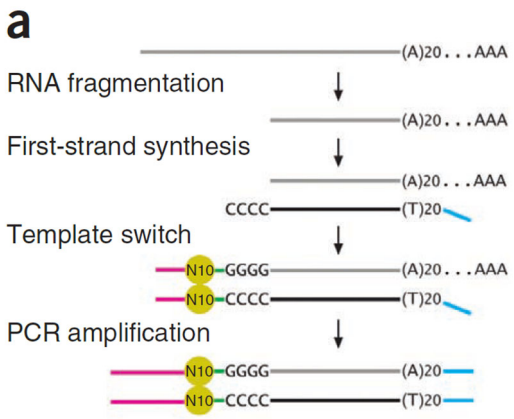
448. Kivioja further explains in Supplementary Table 1 that the green sequence was "an index sequence that was used to enable multiplexed sequencing," EX1012 (Kivioja) at 0014, *i.e.* a sample index or an indexing site as the '241 patent defines it.

n. Claim 14: "The method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length."

449. Kivioja anticipates claim 14.

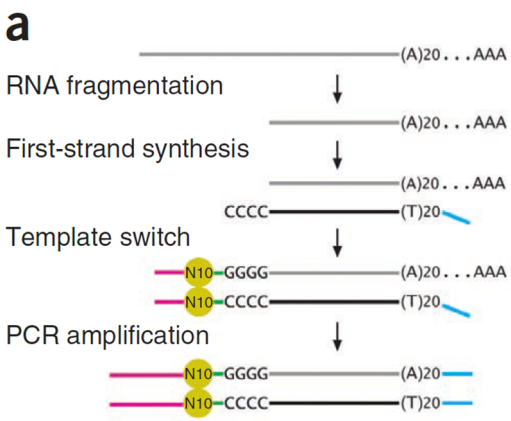
450. As discussed further above in Section VI.C.8.a.2 of my declaration, Kivioja used an identifier site of 10 nucleotides in length in the mRNA proof of concept experiment. The gold "N10," shown in Figure 3a reproduced below, is

described in the Kivioja method as "10-base-pair random label." EX1012 (Kivioja) at 0002. This is further consistent with the sequence of the 5′ adaptor provided in Kivioja Supplementary Table 1. EX1012 (Kivioja) at 0014.



o. Claim 15: "The method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons."

451. Claim 15 is anticipated by Kivioja. As I explain further above in Section VI.C.8.a of my declaration in connection with Claim 1, Kivioja disclosed each of these steps as part of the mRNA sequencing method. Kivioja first obtained an RNA sample from *Drosophila* cells. EX1012 (Kivioja) at 0004. Then, Kivioja fragmented the RNA to obtain RNA fragments. EX1012 (Kivioja) at 0004. Then after transcription to cDNA and adaptor ligation, Kivioja performed an amplification reaction to obtain the amplicons. EX1012 (Kivioja) at 0004. Kivioja showed this reaction in Figure 3b, reproduced below. EX1012 (Kivioja) at 0002.



p. Claim 16: "The method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index."

452. Claim 16 would have been anticipated by and/or obvious based on Kivioja. Claim 16 recites a routine implantation for sample multiplexing. Kivioja anticipates this limitation by disclosing that the adaptor used in the mRNA experiment contained "an index sequence that was used to enable multiplex sequencing." EX1012 (Kivioja) at 0014. As I discuss above in connection with Claims 7 and 17 of the '399 patent in Sections VI.C.6.f and VI.C.6.n above, respectively, the separation of reads by sample index was a standard part of Illumina workflows on the devices used by Kivioja.

453. To the extent Kivioja does not anticipate this claim, at a minimum, this claim would have been obvious to the POSA based on Kivioja. Specifically, the

pooling of samples for multiplexing using sample indexes was routine at the priority date as I discuss above in Section IV.K.2.a. The POSA would have recognized that the simplest implementation of this concept was to pool two samples wherein each have a different sample index. Moreover, separation of the samples was a standard part of Illumina workflows at the priority date as I discuss above in connection with Claims 7 and 17 of the '399 patent in Sections VI.C.6.f and VI.C.6.n above. In view of Kivioja's express disclosure of sample multiplexing, the POSA would have had a reasonable expectation of success in practicing the method of claim 16.

D. The Challenged Claims Of The '399, '357, And '241 Patents Are Anticipated And/Or Obvious Over Bielas.

454. In this section of my declaration I provide the reasoning for my opinion that claims 1–3, 5–11, 14–17 of the '399 patent; claims 1–4, 6–7, and 9–10 of the '357 patent; and claims 1–16 of the '241 patent would have been obvious and/or anticipated by Bielas by November 13, 2013.

455. Bielas was published on August 22, 2013 as International Publication No. WO 2013/123442. EX1013 (Bielas). I understand Bielas is available as prior art to the '399, '357, and '241 patents because it published before the November 13, 2013, priority date of those patents.

1. Bielas Disclosed Methods With All Of The Elements Claimed In The '399, '357, And '241 Patents.

456. Bielas explains that it "relates to compositions and methods for

accurately detecting mutations using sequencing." EX1013 (Bielas) at 0003. As Bielas explains, "[t]he most direct way to detect mutations is by sequencing, but the available sequencing methods are not sensitive enough to detect rare mutations" "since a mutation must be present in as much as 10–25% of a population of molecules to be detected by sequencing." EX1013 (Bielas) at 0003–04. To address this issue, Bielas explains that PCR amplification is generally necessary prior to NGS. EX1013 (Bielas) at 0004. But as Bielas notes, "the amplification of single DNA molecules prior to or during sequencing by PCR and/or bridge amplification suffers from the inherent error rate of polymerases employed for amplification, and spurious mutations generated during amplification may be misidentified as spontaneous mutations from the original (endogenous unamplified) nucleic acid." EX1013 (Bielas) at 0004. Bielas also discusses other sources of error, e.g., damage to the DNA during sample preparation. EX1013 (Bielas) at 0004. In addition to library preparation, Bielas explains that "[m]assively parallel sequencing cannot generally be used to detect rare variants because of the high error associated with the sequencing process—one process using bridge amplification and sequencing by synthesis has shown an error rate that varies from about 0.06% to 1%, which depends on various factors including read length, base-calling algorithms, and the type of variants detected." EX1013 (Bielas) at 0004 (citing Kinde, EX1025).

457. To address the problems known in the prior art, Bielas teaches a method

of preparing a double-stranded nucleic acid molecule library "that includes a plurality of target nucleic acid molecules and a plurality of random cyphers." EX1013 (Bielas) at 0004–05. Bielas explains that a Cypher, which Bielas also refers to as a "random cypher," "cypher," "barcode," or "identifier tag," EX1013 (Bielas) at 0010, is a random molecular tag used to uniquely label nucleic acid molecules in a sample, e.g., EX1013 (Bielas) at 0008, 0026, 0038. Labeling a sample of nucleic acid molecules with Cyphers allows "a library of molecules" to "be individually reconstructed." EX1013 (Bielas) at 0008. As Bielas explains, Cyphers "allow a person of ordinary skill in the art to more accurately distinguish true mutations (i.e., naturally arising *in vivo* mutations) of a nucleic acid molecule from artifact 'mutations' (i.e., *ex vivo* mutations or errors) of a nucleic acid molecule that may arise for various reasons, such as a downstream amplification error, a sequencing error, or physical or chemical damage." EX1013 (Bielas) at 0007.

458. Bielas explains that "the present disclosure provides a method for obtaining a nucleic acid sequence or accurately detecting a true mutation in a nucleic acid molecule by amplifying each strand of the aforementioned double-stranded nucleic acid library wherein a plurality of target nucleic acid molecules and plurality of random cyphers are amplified, and sequencing each strand of the plurality of target nucleic acid molecules and plurality of random cyphers." EX1013 (Bielas) at 0005.

459. Thus, Bielas teaches a method of using random molecular tags to identify PCR duplicates of original molecules and then deduplicate those PCR copies in order to determine, for example, whether a particular mutation is present in the original fragment (a "true mutation") or was created during PCR amplification (an "artifact mutation").

a. Bielas Example 1

460. Example 1 and Example 3 in Bielas provide further details and proof of concept. In Example 1, Bielas applies the method to a "tumor genomic library." EX1013 (Bielas) at 0026–27. In Example 1, Bielas uses Cyphers "[t]o examine whether malignant cells exhibit a mutator phenotype resulting in the generation of random mutations throughout the genome." EX1013 (Bielas) at 0026. Bielas first obtained "genomic DNA from patient-matched normal and tumor tissues" and then fragmented the DNA to prepare the sample for ligation to adaptors. EX1013 (Bielas) at 0026. Bielas ligated the DNA fragments into a vector, or circularized DNA, containing adaptors: "The end-repaired DNA fragments are ligated into the *Sma*I site of the library of dual cypher vectors shown in Figure 2 to generate a target genomic library." EX1013 (Bielas) at 0026. I have reproduced Figure 2 below. As the figure shows, the *Sma*I site corresponds to "RE3" in Figure 2 and this is the position where the target DNA fragments were ligated. EX1013 (Bielas) at 0041.

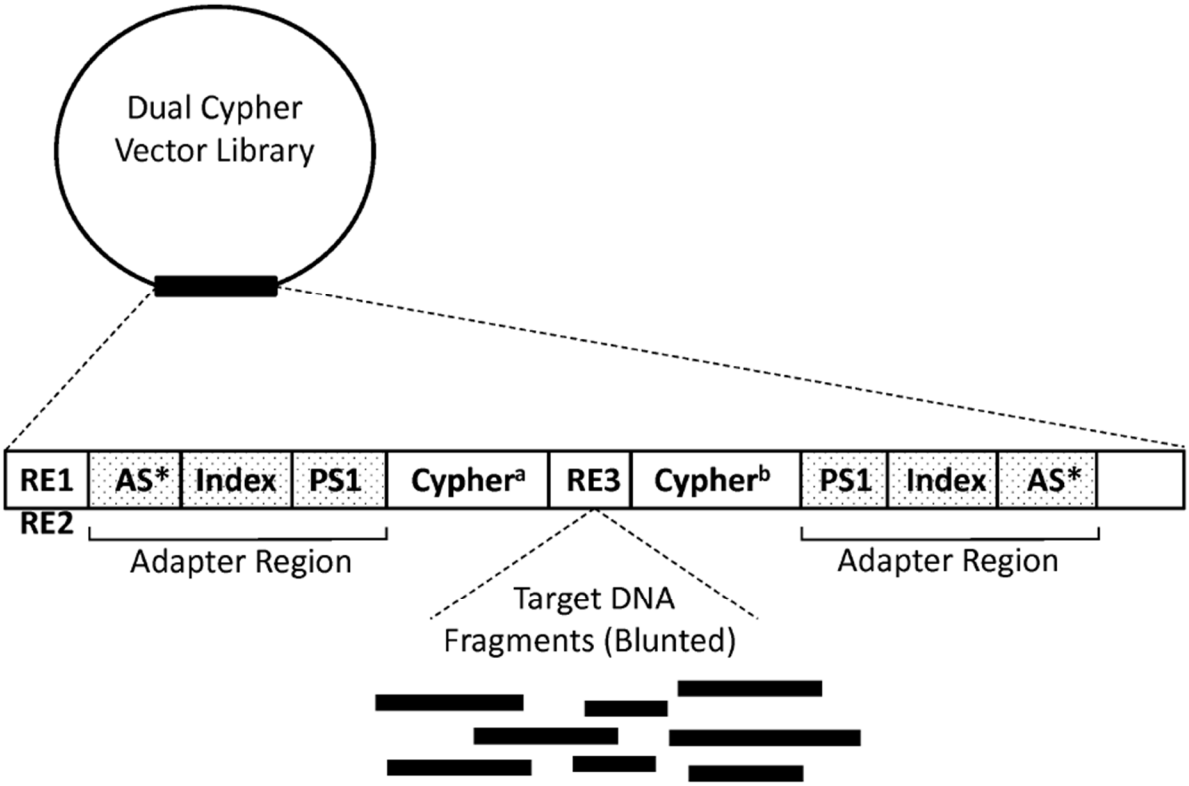
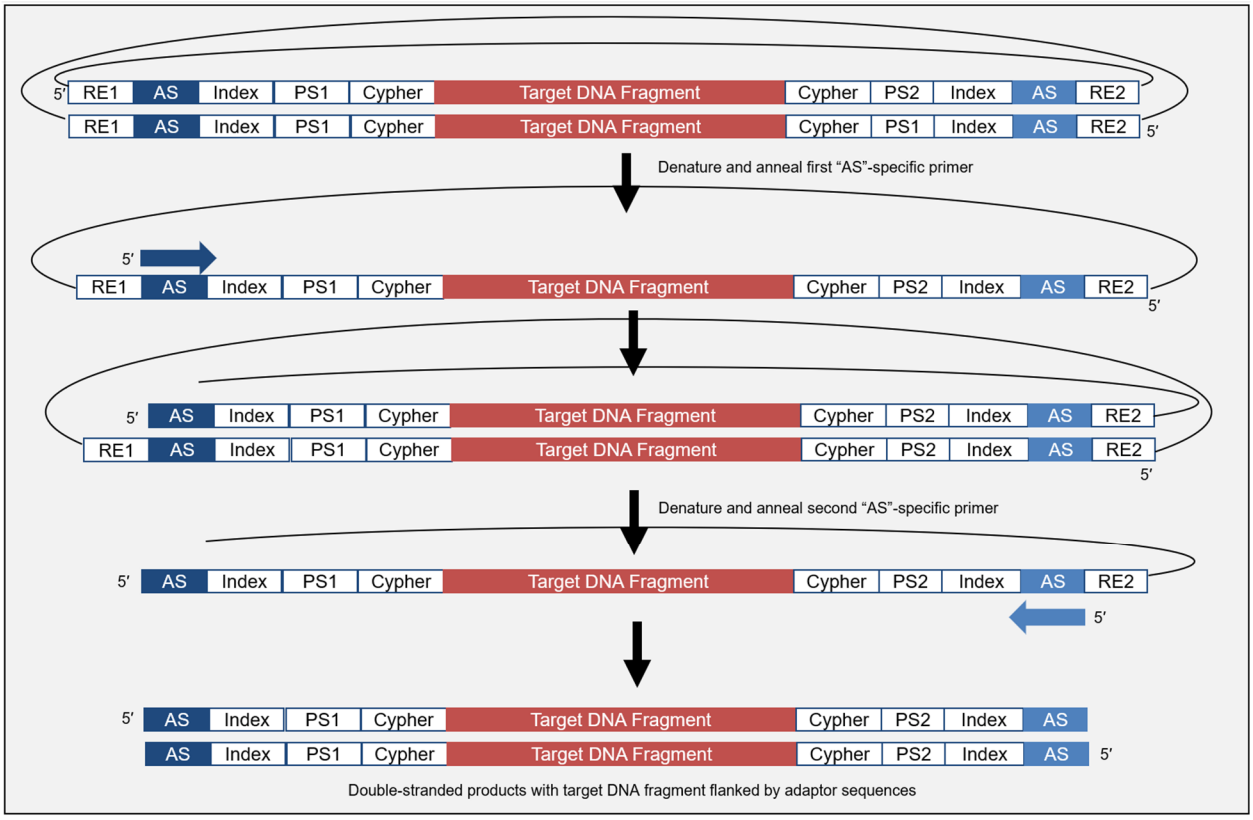


Fig. 2

461. The *Sma*I or RE3 site is a restriction endonuclease recognition site that allowed Bielas to cut the vector with a restriction enzyme and insert the DNA fragments from the sample. EX1013 (Bielas) at 0023. The 3′ end of the circular vector is ligated to the 5′ end of the fragment and the 5′ end of the vector to the 3′ end of the fragment. EX1013 (Bielas) at 0026–27.

462. After ligation, Bielas explained that "[t]he target genomic library fragments are amplified" and "the amplification is performed using sense strand and anti-sense strand primers that anneal to a sequence located within the adapter region." EX1013 (Bielas) at 0026–27. As Bielas explained, "the primer will anneal

to a sequence upstream of the AS" region. EX1013 (Bielas) at 0027. I have provided a figure showing that PCR reaction below. The result of that PCR reaction is to amplify only the adaptor-ligated regions out of the circular vector, and to create a library of double-stranded DNA molecules with the fragments between the two adaptors, as shown at the bottom of the figure. By using this PCR reaction, the Bielas authors achieved the same result as cutting out the adaptor-ligated fragment region of the circularized vector without needing to amplify the entire circularized vector. EX1013 (Bielas) at 0026-27, 0041.



463. The resulting adaptor-ligated amplified products are a library of double-stranded DNA molecules with the structure: AS-Index-PS1-Cypher-

Fragment-Cypher-PS2-Index-AS. EX1013 (Bielas) at 0026–27, 0041. That structure corresponds to the structure shown in Bielas's Figure 2. EX1013 (Bielas) at 0041.

464. On the 5′ end of the fragment, the ligated adaptor has the structure of AS-Index-PS1-Cypher. EX1013 (Bielas) at 0026–27, 0041. Bielas explains that "AS" is "an adapter sequence." EX1013 (Bielas) at 0022. Based on that teaching, the position of the AS sequence in the adapter-ligated fragment, and Bielas's teaching that the library was sequenced on an Illumina Genome Analyzer II, the POSA would understand that AS was a common Illumina adaptor sequence that allows the adaptor-ligated fragment to bind to the fixed sequences on the surface of an Illumina device. EX1013 (Bielas) at 0022, 0026–27, 0038. As I discuss above in Section IV.K.2.a ¶¶ 110-111 of my declaration, the POSA further would have understood that for certain Illumina devices, the AS sequence also functions as the indexing primer binding site.

465. Bielas further explains that "Index" is a sample index used for sample pooling "if multiplex sequencing is desired." EX1013 (Bielas) at 0022, 0027. As Bielas explains, in certain embodiments, including Example 1, "a unique index sequence (comprising a length ranging from about 4 nucleotides to about 25 nucleotides) specific for a particular sample is included with each dual cypher vector library. For example, if ten different samples are being pooled in preparation for

multiplex sequencing, then ten different index sequences will be used such that ten dual cypher vector libraries are used in which each library has a single, unique index sequence identifier (but each library has a plurality of random cyphers)." EX1013 (Bielas) at 0015–16. The POSA would understand that this teaching regarding sample indexes was consistent with the well-known use of sample indexes in NGS, and specifically on Illumina devices, at the priority date.

466. "PS1" is a primer binding site for the Illumina SBS primer that reads the fragment sequence. EX1013 (Bielas) at 0012, 0022–23, 0027. Bielas explains that PS is a "nucleic acid molecule priming site" and "[a] nucleic acid molecule primer that is complementary to a PS included in a library of the present disclosure can be used to initiate a sequencing reaction." EX1013 (Bielas) at 0012. In the context of Example 1, which used an Illumina device, the POSA would understand PS was the binding site for the forward read primer in the SBS reaction on the Illumina device.

467. In addition to being the forward read primer binding site, for certain Illumina devices, including the Genome Analyzer II that Bielas used in Example 1, the PS1 sequence also functions as the indexing primer binding site as I discuss above in Section IV.K.2.a ¶¶ 110-111 of my declaration. EX1055 (Kircher) at 0002.

468. Finally, the "Cypher" is a random molecular tag sequence. EX1013 (Bielas) at 0016, 0026, 0038. Bielas explains that "[a]s used herein, the term

'random cypher' or 'cypher' or 'barcode' or 'identifier tag' and variants thereof are used interchangeably." EX1013 (Bielas) at 0010. Bielas explains that "the present disclosure, in a further aspect, provides methods for identifying mutations present before amplification or sequencing of a double-stranded nucleic acid library wherein the target molecules include a single double-stranded cypher or dual cyphers (i.e., barcodes or identifier tags), one on each end, so that sequencing each complementary strand can be connected back to the original molecule." EX1013 (Bielas) at 0009.

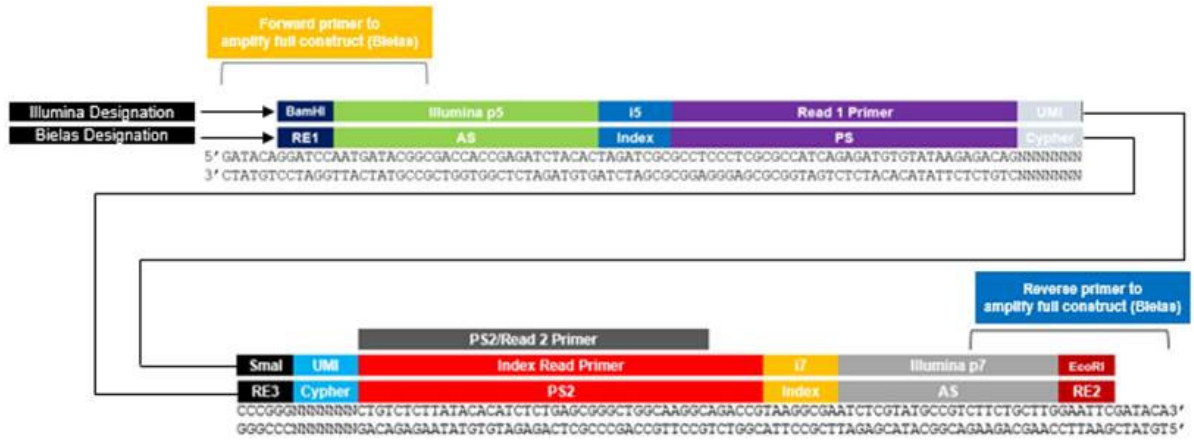
469. Bielas then explains that "[t]he sequencing of the library described above will be performed using, for example, an Illumina Genome Analyzer II sequencing instrument as specified by the manufacturer." EX1013 (Bielas) at 0027. The result of this NGS sequencing run would have been a population of demultiplexed sequencing reads, separated into FASTQ files. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35.

470. After sequencing, Bielas explains that "the unique cypher tags are used to computationally deconvolute the sequencing data and map all sequence reads to single molecules (i.e., distinguish PCR and sequencing errors from real mutations." EX1013 (Bielas) at 0027. This process necessarily required the identification of duplicate sequencing reads using the Cypher sequence and the sequence of the DNA fragment. As explained, one piece of information Bielas used to deconvolute the sequencing information was the Cypher tag. EX1013 (Bielas) at 0007-08. But in

addition as part of the same process, Bielas used the sequence of the DNA fragment for "sequence alignment," using Illumina's Eland pipeline software package. EX1013 (Bielas) at 0027. To align a sequenced DNA fragment from a sample of human origin, as in Example 1, the fragment is mapped to its corresponding position on the then-known human genome. Sequence alignment allows the determination of whether any mutations are present in the sequenced DNA fragment because mutations in the fragment would not be present in the then-known human genome reference sequence. Thus, as part of this process, Bielas used both the Cypher sequence and the fragment sequence to deconvolute, or in other words, to de-deduplicate, the sequencing data.

b. Bielas Example 3

471. In Example 3, Bielas used a method similar to the one used in Example 1. In Example 3, Bielas describes a method for high-resolution detection of mutations in human TP53, a tumor suppressor that is mutated in certain cancers. EX1013 (Bielas) at 0028–32. Bielas first constructed an adaptor-vector library, much like the one in Example 1, EX1013 (Bielas) at 0028–29, and then isolated and amplified a portion of the TP53 gene (exon 4) from human subjects for ligation to the adaptor-vectors. EX1013 (Bielas) at 0030–31. For Example 3, Bielas provided the sequence of the adaptors, which I have annotated below and in EX1075 (Bielas Annotated Figure). EX1013 (Bielas) at 0028–29.



472. The annotations I provide in this figure are based on the publicly-available prior art information about the standard sequences used in Illumina NGS devices. Illumina published the sequences used in its NGS devices to anneal to the chip surface, including the P5 adaptor sequence, and for the SBS reaction primers to bind, including indexing read primers. EX1056 (Illumina Letter) at 0003. Illumina likewise published the sequences of its conventional sample indexes, including its so-called "i5" index. EX1056 (Illumina Letter); EX1057 (Nextera Prep Guide). Other references in the prior art reported these sequences routinely, reflecting the fact that they were publicly available and standard knowledge to the POSA at the priority date. EX1053 (Syed); EX1054 (Loeb); EX1055 (Kircher). Likewise, groups like Bielas would incorporate these standard Illumina sequences into their library preparation methods to ensure compatibility with Illumina devices. The POSA would have immediately recognized from the disclosure of Bielas that the sequence of the Example 3 adaptor contained these standard Illumina sequences.

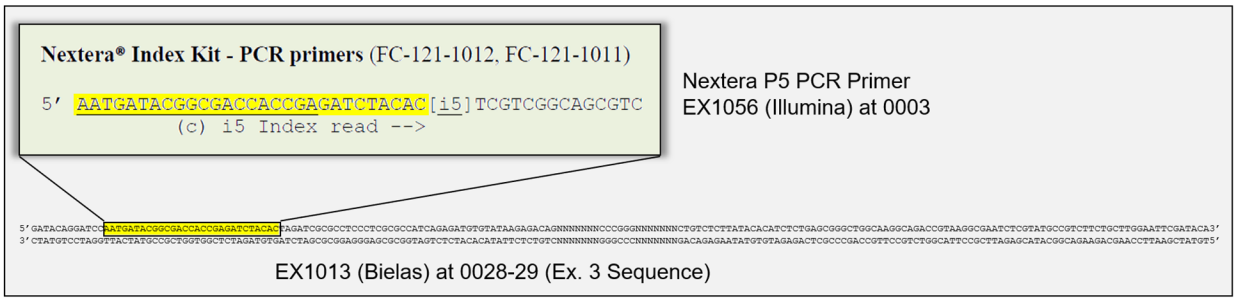
473. The adaptor features used in Example 3 were similar to those used in Example 1: AS-Index-PS1-Cypher-Fragment-Cypher-PS2-Index-AS. EX1013 (Bielas) at 0028–29. On the 5′ end of the DNA fragment, the sequence of the adaptor, like in Example 1, is AS-Index-PS1-Cypher. EX1013 (Bielas) at 0028.

474. As I have shown in the figure above, the POSA would have recognized that the adaptor corresponded to the features described in Bielas and that these features corresponded to standard Illumina adaptor sequences. Bielas's "AS" sequence is the Illumina P5 sequence, which as described above functions as both the region that anneals to the fixed sequences bound to the Illumina flow cell and as the indexing primer binding site for certain Illumina devices, such as the MiSeq, which Bielas used in Example 3. EX1013 (Bielas) at 0028–29, 0031; EX1053 (Syed) at 0004. The Illumina P5/AS sequence would have been known to the POSA based on the disclosures in the prior art. The prior art clearly showed that the Illumina P5 adapter sequence was 5′-AATGATACGGCGACCACCGAGATCTACAC-3′. EX1056 (Illumina Letter) at 0003, 0005, 0006, 0007, 0010, 0011. The prior art disclosed this sequence as the i5 index read primer binding site (i.e., P5) for use in the Illumina Nextera® Sample preparation, EX1056 (Illumina Letter) at 0003. This P5 primer sequence is disclosed, for example, in the below figure I have reproduced in excerpted form from EX1056 (Illumina Letter) at 0003.

Nextera® Index Kit - PCR primers (FC-121-1012, FC-121-1011)

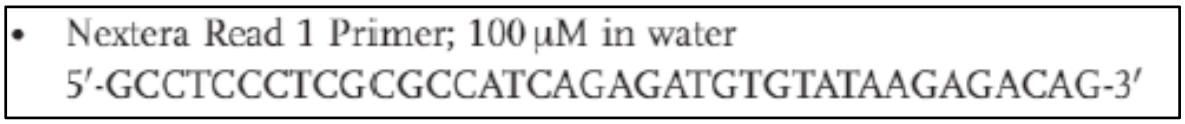
5′ AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC
(c) i5 Index read -->

475. Illumina Nextera® is a method used to produce samples for sequencing on Illumina sequencers, including the GAIIx system, used in Example 1 of Bielas, and the MiSeq system, used in Example 3 of Bielas. EX1057 (Nextera Prep Guide) at 0005, 0039–43; EX1013 (Bielas) at 0026–32. The sequence of the Illumina P5, which is disclosed in Illumina as the stretch of nucleotides 5′ upstream of the i5 site in the above excerpted oligonucleotide, corresponds to nucleotides 13 to 41 of SEQ ID NO:2 in Example 3 of Bielas, as shown in the annotated figure below, also available as EX1076 (Bielas Illumina P5). EX1013 (Bielas) at 0028–29.

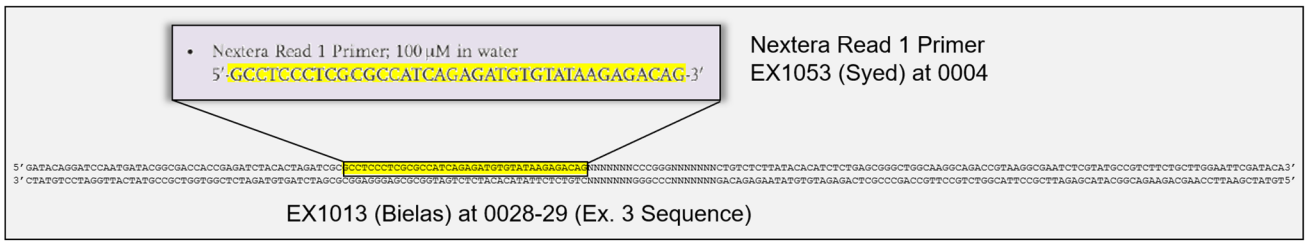


476. "PS1" is the sequencing primer binding site where the Illumina SBS primer binds to read the fragment sequence, specifically the Illumina Read 1 primer site. EX1013 (Bielas) at 0012, 0022–23, 0028; EX1053 (Syed) at 0004. The Read 1 Primer/PS sequence would have been known to a POSA based on the disclosures in the prior art. The prior art clearly showed that the Read 1 Primer/PS sequence

was 5′-GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3′, and was used as a Read 1 Primer for Nextera sample preparation. EX1053 (Syed) at 0004); EX1008 (Caruccio) at 0005. This Read 1 Primer/PS sequence is disclosed, for example, in the below figure I have reproduced in excerpted form from EX1053 (Syed) at 0004; EX1008 (Caruccio) at 0005.



477. As discussed above, Illumina Nextera® is a method used to produce samples for sequencing on Illumina sequencers, including the GAIIx system, used in Example 1 of Bielas, and the MiSeq system, used in Example 3 of Bielas. EX1057 (Nextera Prep Guide) at 0005, 0039–43. The sequence of the Illumina Read 1 primer, which is disclosed in Syed and Caruccio, corresponds to nucleotides 50 to 87 of SEQ ID NO:2 in Example 3 of Bielas. EX1013 (Bielas) at 0028–29. I illustrate the location of the Illumina Read 1 primer within the full Bielas sequence in the annotated figure below, also available as EX1077 (Bielas Read 1 Primer).

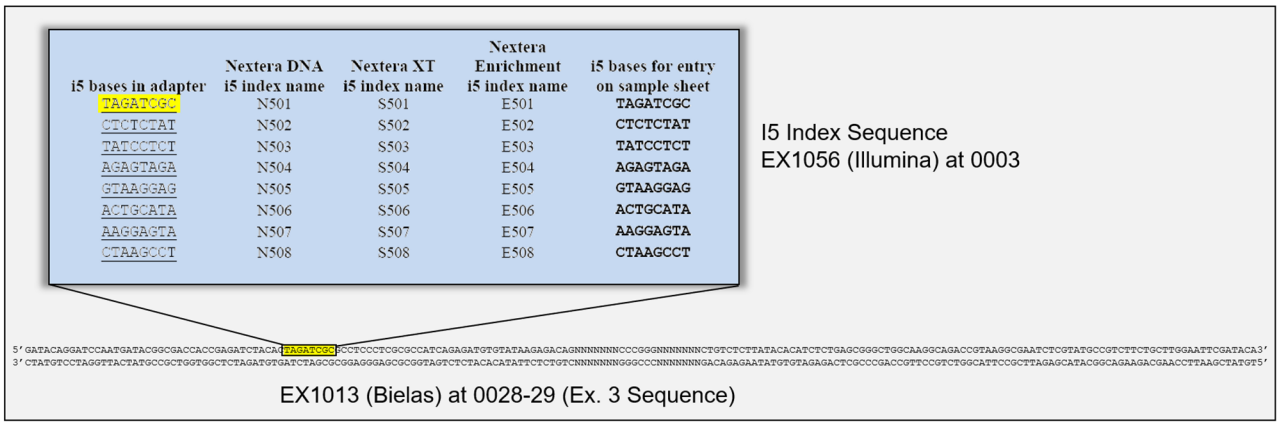


478. In addition, the Bielas adaptor included a known Illumina sample index, a so-called i5 index, TAGATCGC. EX1056 (Illumina Letter) at 0003; EX1057

(Nextera Prep Guide) at 0044. This sequence is disclosed, for example, as the first i5 index in the below figure I have reproduced in excerpted form from EX1056 (Illumina Letter) at 0003.

i5 bases in adapter	Nextera DNA i5 index name	Nextera XT i5 index name	Nextera Enrichment i5 index name	i5 bases for entry on sample sheet
TAGATCGC	N501	S501	E501	**TAGATCGC**
CTCTCTAT	N502	S502	E502	**CTCTCTAT**
TATCCTCT	N503	S503	E503	**TATCCTCT**
AGAGTAGA	N504	S504	E504	**AGAGTAGA**
GTAAGGAG	N505	S505	E505	**GTAAGGAG**
ACTGCATA	N506	S506	E506	**ACTGCATA**
AAGGAGTA	N507	S507	E507	**AAGGAGTA**
CTAAGCCT	N508	S508	E508	**CTAAGCCT**

479. A POSA would have recognized this i5 sequence, which was known in the art at least as early as 2012. EX1056 (Illumina Letter) at 0003; EX1057 (Nextera Prep Guide) at 0044. The sequence of the i5 index, which is disclosed in Illumina and the Nextera Prep Guide, corresponds to nucleotides 42 to 49 of SEQ ID NO:2 in Example 3 of Bielas. EX1013 (Bielas) at 0028 –29. I demonstrate the location of this i5 index within the full Bielas sequence in the annotated figure below, also available as EX1078 (Bielas i5 Index).

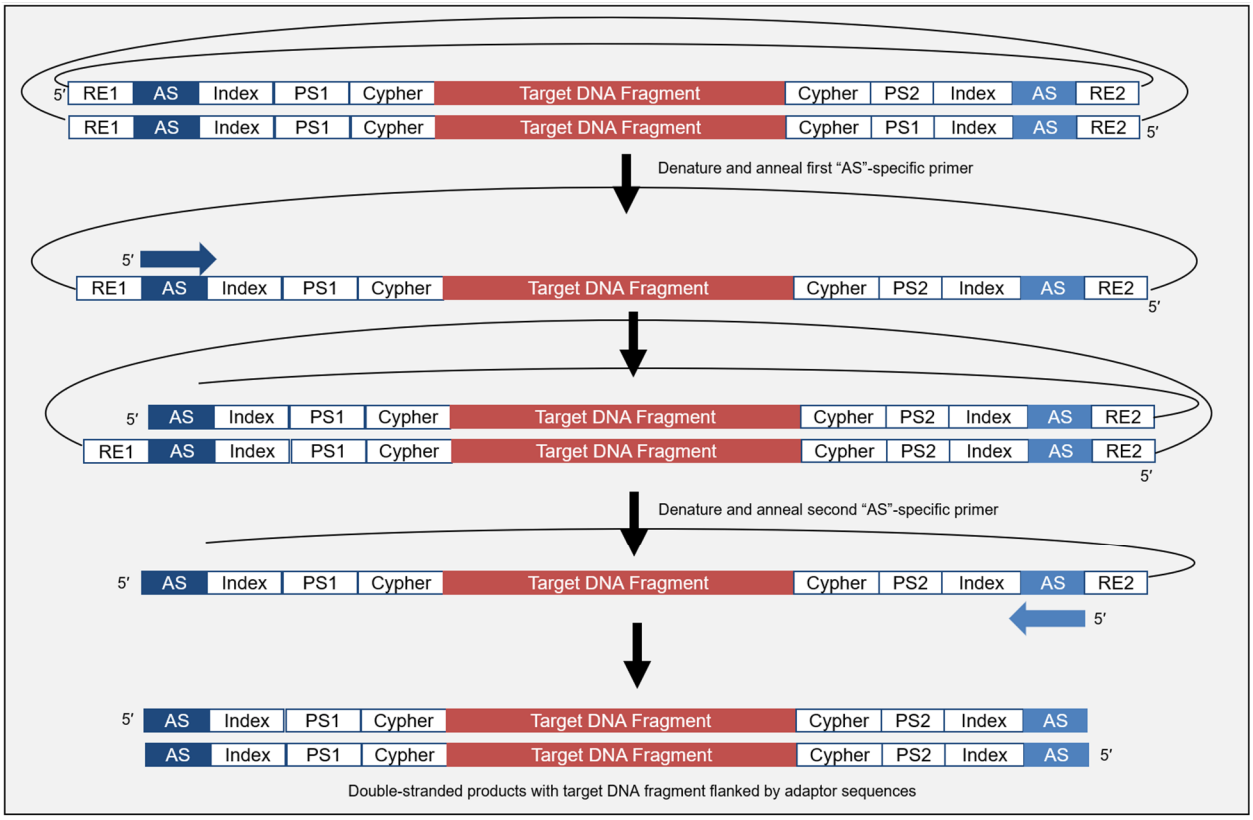


480. The "Cypher" is a 7 nucleotide random molecular tag, which Bielas represented in the fragment as series of "N"s—the standard notation for random bases. EX1013 (Bielas) at 0016, 0026, 0028–29, 0038. As in Example 1, Bielas explains "[t]he Cypher Seq vector library was digested with *Sma*I" and purified. EX1013 (Bielas) at 0030. "Blunt-end ligations of the vector and TP53 exon 4 DNA were then carried out in 20 µl reactions using T4 DNA Ligase HC." EX1013 (Bielas) at 0031.

481. As in Example 1, Bielas then amplified the adaptor-fragment regions out of the vector by PCR to create a double-stranded library of linear adaptor-ligated fragments. Specifically, Bielas explains "[t]he sequencing construction containing the Illumina adapters, barcodes, and TP53 DNA was then amplified using 10 cycles of PCR and primers designed against the adapter ends (forward primer: AATGATACGGCGACCACCGA, SEQ ID NO.7; and reverse primer: CAAGCAGAAGACGGCATACGA, SEQ ID NO.8)." I provide two figures below

showing this PCR reaction, one with sequence information and one in simplified form.


RE1
AS (Illumina p5)
i5 Index
PS (Read 1 Primer)
Cypher
RE3
Cypher
PS2 (Index Read Primer)
i7 Index
AS (Illumina p7)
RE2
Denature and anneal forward and reverse primers
AGCATACGGCAGAAGACGAAC 5'
Reverse primer
Forward primer
5' AATGATACGGCGA
Resulting double-stranded products



482. Bielas then explains that following purification the resulting library was sequenced on an Illumina MiSeq device "as instructed by the manufacturer's protocol." EX1013 (Bielas) at 0031.

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.

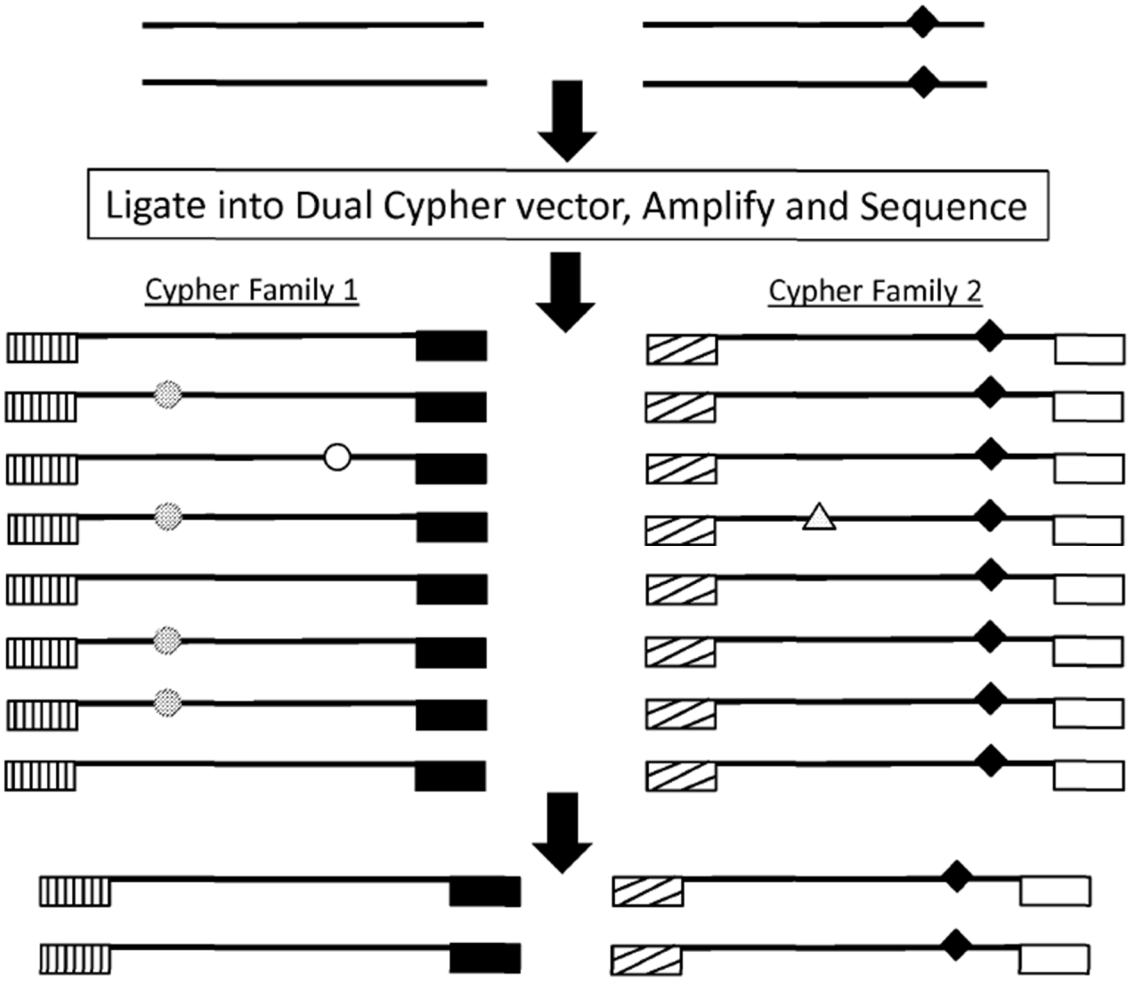


Fig. 3B

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.

485. At the top of Figure3B, there are two original double stranded nucleic acid fragments from a sample. EX1013 (Bielas) at 0028–32, 0043. As described above, those nucleic acid fragments are ligated to adaptor-vectors, wherein the adaptor contains a Cypher/identifier site, sample index, indexing primer binding site, and sequencing primer binding site. EX1013 (Bielas) at 0028–29, 0043. Those adaptors are shown as boxes at either end of the nucleic acid fragments in Figure 3B. EX1013 (Bielas) at 0043. Figure 3B then explains that the adaptor-ligated nucleic acid fragments are PCR amplified to create copies of the original fragments. EX1013 (Bielas) at 0043. Finally, at the bottom of Figure 3B, representing the end of the process, Bielas shows that PCR duplicates of original molecules were identified and copies were deduplicated into a single sequencing read. EX1013 (Bielas) at 0043, 0031–32.

486. As I explain further in the annotated version of the figure below, Figure 3B depicts several steps. First, it shows two fragments from the original sample, one with a black diamond mutation. Then, those fragments are duplicated via PCR to create additional copies. Finally, the duplicate reads are de-duplicated back to a single fragment. As part of that deduplication process, Bielas considered the Cypher sequences shown in the boxes on either end of the fragment and the fragment sequence. This is evidenced by the fact that in the right hand family of fragments, the true mutation is reflected in the consensus sequence but the white triangle false

mutation is removed. To perform that analysis, Bielas necessarily aligned the fragments to the reference genome so as to ascertain the presence of the true mutation, and compared the observed sequences to one another to establish that the white triangle was a false mutation that could be removed from the sequencing data. This is demonstrated in Figure 3B of EX1013 (Bielas), which I have reproduced and annotated below. EX1013 (Bielas) at 0043; 0031–32.

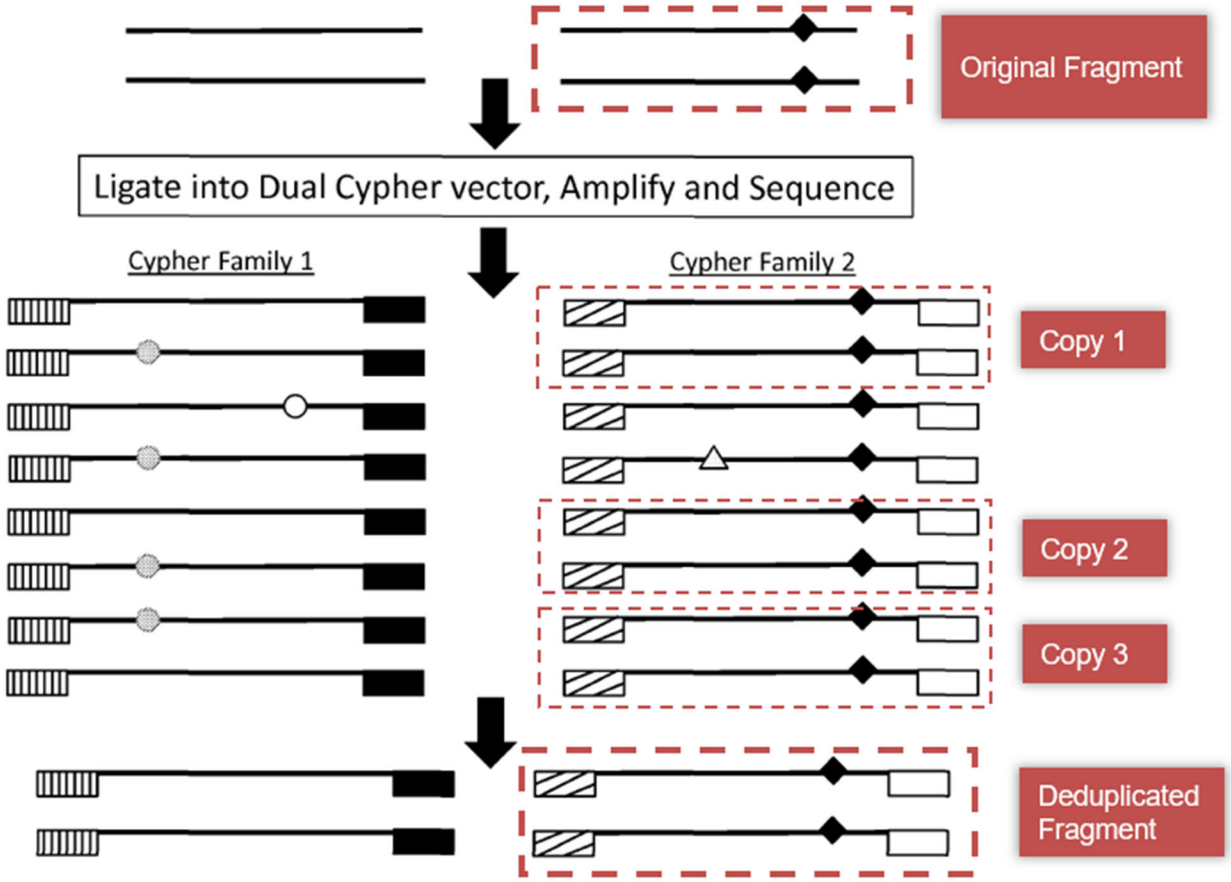


Fig. 3B

2. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The ’399 Patent

487. Bielas teaches every element of the method claimed in claims 1–3, 5–11, and 14–17 of the ’399 patent and thus anticipates the claims and/or renders the claims obvious. Specifically, Bielas teaches a method using adaptors containing all of the elements recited in the claims of the ’399 patent, including a sample index, an identifier site (which Bielas refers to as a “Cypher”), a sample index primer binding site, and a target sequence primer binding site. EX1013 (Bielas) at 0041, 0038. Bielas further teaches using the Cypher and sequence of the nucleic acid fragment to identify PCR copies of original fragments after an NGS read, which necessarily requires identifying a duplicate sequencing read. EX1013 (Bielas) at 0026-32.

a. Bielas Anticipates Claim 1.

488. In my opinion, for the reasons stated in Section VI.D.1 of my declaration and the following reasons, Bielas anticipates claim 1 of the ’399 patent.

1) “A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:”

489. Bielas discloses this limitation. As described above in Section VI.D.1, Bielas teaches a method of identifying mutations and determining whether such mutations are “true mutations” or “artifact mutations.” EX1013 (Bielas) at 0007.

490. As shown in Example 1 and Example 3, to accomplish this, Bielas uses a combination of the sequence of the Cypher and the sequence of the nucleic acid

fragment. EX1013 (Bielas) at 0026–32. In Example 1, Bielas describes using the Cypher and "sequence alignment" to the reference genome to "computationally deconvolute" the data to "distinguish PCR and sequencing errors from real mutations." EX1013 (Bielas) at 0026–27. In Example 3, and as diagrammatically shown in Figure 3B, Bielas describes identifying true mutations from a process of using the Cypher and sequence of the nucleic acid fragment to identify duplicate sequencing reads. EX1013 (Bielas) at 0028–29, 0031–32. Specifically, Bielas used the Cypher and sequence of the nucleic acid fragment to identify a consensus sequence after which "errors introduced during library preparation and during sequencing were computationally eliminated." EX1013 (Bielas) at 0031–32.

491. The POSA would have known that to "deconvolute" sequences, or to "computationally eliminate" fragments with PCR or sequencing error, necessarily required identification of duplicate sequencing reads as a predicate step.

2) "a) ligating an adapter to a 5′ end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples,"

492. Bielas discloses this limitation. As I describe in detail above in Section VI.D.1, in both Examples 1 and 3 the Bielas authors ligated the fragments to be sequenced into a circular vector with adaptors on either end of the fragment, EX1013 (Bielas) at 0026–29. As shown in the below annotation of Bielas Figure 2, adaptors are ligated onto either end of the fragment. On the 5′ end, the adaptor has the

sequence AS-Index-PS1-Cypher. As I discuss above in Section VI.D.1, Bielas taught that these adaptor-ligated fragments could later be amplified or enzyme digested out of the circular vector. This is shown below in Figure 2 of EX1013 (Bielas), which I have reproduced and annotated below. EX1013 (Bielas) at 0043, 0026–29.

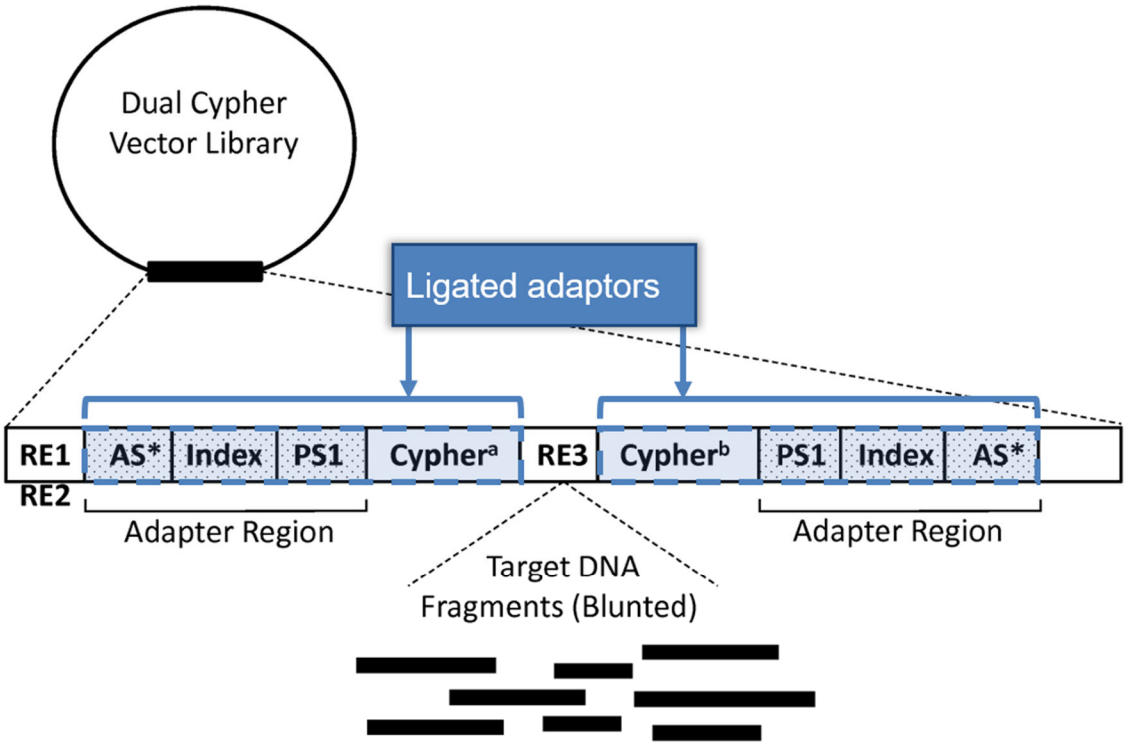


Fig. 2

493. Bielas further teaches that the adaptors can be ligated to nucleic acid fragments from more than one sample. Indeed, Bielas included a sample index ("Index") in the adaptor in both Example 1 and Example 3. EX1013 (Bielas) at 0027–28, 0043. Bielas further explained that the sample index was for multiplexed

sequencing. EX1013 (Bielas) at 0027. The POSA would have understood that the inclusion of a sample index expressly contemplated pooling multiple samples and that the methods used in Bielas were suitable for multiplexing.

i. "wherein the adapter comprises: (i) an indexing primer binding site;"

494. Bielas discloses this limitation. In Example 1, Bielas sequenced the adaptor-DNA fragments on an Illumina Genome Analyzer II. EX1013 (Bielas) at 0027. As I discuss above in Section VI.D.1.a, for sequencing on an Illumina Genome Analyzer II, the POSA would have known that the PS1 sequence functions as the indexing primer binding site. In Example 3, Bielas sequenced the adaptor-DNA fragments on an Illumina MiSeq. EX1013 (Bielas) at 0031. As I discuss above in Section VI.D.1.b, for sequencing on an Illumina MiSeq, the POSA would have known that the "AS" sequence functions as the indexing primer binding site. Specifically, the sequence 5′-AATGATACGGCGACCACCGAGATCTACAC-3′ in the Example 3 adaptor was the indexing primer binding site. EX1056 (Illumina Letter) at 0003.

ii. "(ii) an indexing site;"

495. Bielas discloses this limitation. In Example 1, Bielas cites to Figure 2 and explains that the adaptor ligated to the 5′ end of the nucleic acid fragment contains a sample index "if multiplex sequencing is desired," which is shown in

EX1013 (Bielas) Figure 2, which I have reproduced below (referred to as the "Index"). EX1013 (Bielas) at 0027, 0041.

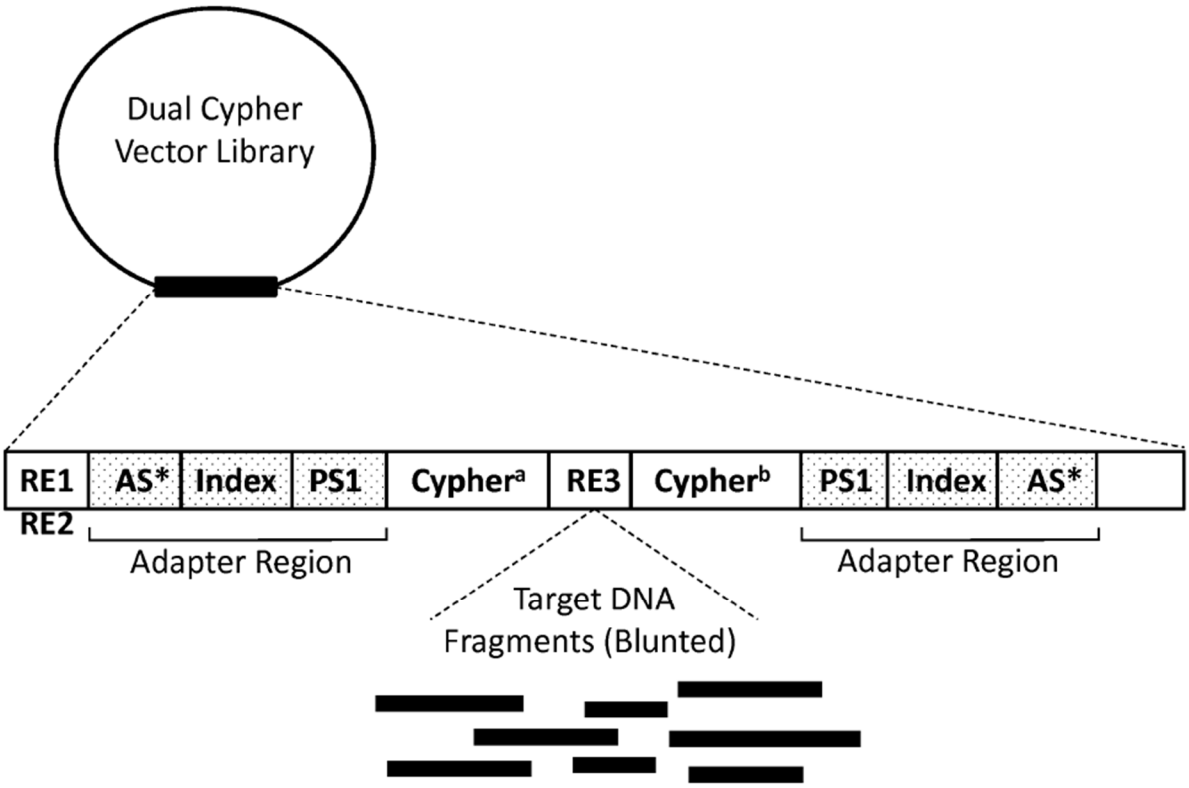
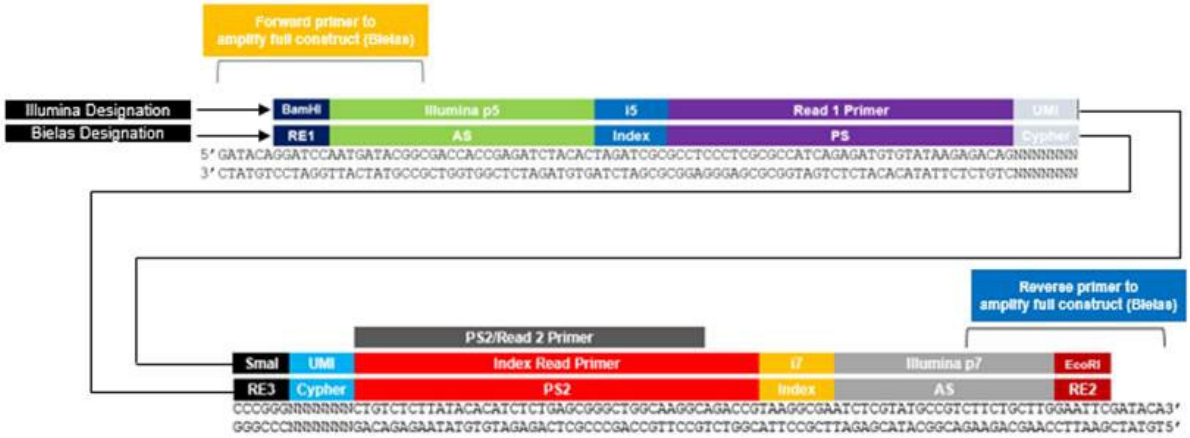


Fig. 2

496. In Example 3, Bielas provides the adaptor sequence. EX1013 (Bielas) at 0028-29. That adaptor sequence, as shown in the figure below and EX1075 (Bielas Annotated Figure), contains a sample index on the 5′ end of the nucleic acid fragment, specifically an Illumina i5 index, TAGATCGC. EX1013 (Bielas) at 0028-29.



497. As I discuss further above in Section VI.D.1.b, the POSA would have recognized this as a standard Illumina i5 index and would have understood that it would only be included for purposes of sample multiplexing.

iii. "(iii) an identifier site consisting of between 1 and 8 nucleotides; and

498. Bielas discloses this limitation. As described further above, Bielas teaches the use of a Cypher—or "identifier site" in the '399 patent's terms—that is between 1 and 8 nucleotides and is incorporated into an adaptor. EX1013 (Bielas) at 0026, 0028–29.

499. In Example 1, Bielas explains that the adaptor structure used is shown in Figure 2. EX1013 (Bielas) at 0027. As described, the adaptor contains a Cypher and is ligated to the 5′ end of the nucleic acid fragment. EX1013 (Bielas) at 0041. Bielas also explains that the Cypher can "be located at one or both ends of the target molecule," which necessarily means it can be located on the 5′ end of the nucleic

acid fragment. EX1013 (Bielas) at 0010.

500. Bielas teaches that the cypher can be specifically "about" 6, "about" 7, or "about" 8 nucleotides (among other options). EX1013 (Bielas) at 0005. This disclosure of Cyphers of 6, 7, or 8 nucleotides anticipates the claimed range of 1 to 8 nucleotides.

501. Moreover, Bielas teaches that the X^a cypher on the 5′ side of the fragment can be specifically 6, 7, or 8 nucleotides. EX1013 (Bielas) at 0015. These express teachings of 6, 7, and 8 nucleotides in the Cypher on the 5′ side of the fragment anticipate the claimed range of 1 to 8 nucleotides.

502. Bielas further discloses a range of "about 6 nucleotides to about 8 nucleotides." EX1013 (Bielas) at 0014. This range falls fully within the claimed range and thus anticipates it.

503. On page 0022, Bielas discusses a similar construct to that shown in Figure 2: 5′-RE1-AS-PS1-X^a-Y-X^b-PS1-AS-RE2-3′. Bielas explains that the 5′ Cypher, X^a, can be about 7 to about 9 nucleotides. EX1013 (Bielas) at 0022. Bielas discusses another similar construction on page 0023 and discloses the same range of 7 to 9 for the 5′ Cypher. EX1013 (Bielas) at 0023. This range overlaps with the claimed range of 1 to 8. The POSA would understand there to be no functional difference between the claimed range and the specific values and ranges disclosed in Bielas. Notably, I do not understand the '399 patent to disclose any particular

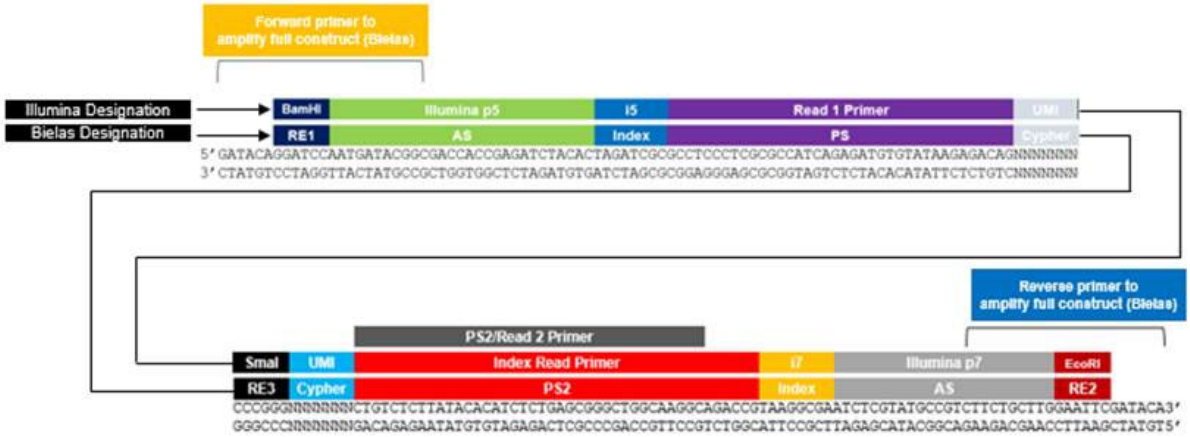
benefit of the range of 1 to 8 compared to the values of 6, 7, or 8, or the range of 7 to 9 disclosed in Bielas.

504. In Example 3, Bielas provides the adaptor structure and specifies a 7-nuclotide Cypher on the 5′ end of the nucleic acid fragment. EX1013 (Bielas) at 0028-29. Bielas identifies the Cypher as "NNNNNNN," EX1013 (Bielas) at 0028-29, which, as explained, the POSA would have understood to be the standard notation for random nucleotides. Bielas's teaching of a 7-nucleotide Cypher in Example 3 anticipates the claimed range of identifier sites of 1–8 nucleotides.

iv. "(iv) a target sequence primer binding site;"

505. Bielas discloses this limitation. As described above in Section VI.D.1, in Example 1 and Example 3, Bielas used an adaptor with the elements AS-Index-PS1-Cyhpher, where "PS1" is the target sequence primer binding site. EX1013 (Bielas) at 0027–29, 0041. For Example 1, where the adaptor structure is disclosed in Figure 2, EX1013 (Bielas) at 0041, sequencing was conducted on an Illumina Genome Analyzer II, EX1013 (Bielas) at 0027; therefore, the POSA would have known that "PS1" was a standard Illumina sequencing primer binding site suitable for use on a Genome Analyzer II. EX1053 (Syed) at 0004; EX1008 (Caruccio) at 0005. For Example 3, Bielas provided the sequence of the adaptor, which, as shown in the annotated figure below, contained a "PS1" sequence corresponding to the

Illumina Nextera read 1 primer. EX1013 (Bielas) at 0028-29; EX1053 (Syed) at 0004; EX1008 (Caruccio) at 0005.



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

506. Bielas discloses this limitation. In Example 1 and Example 3, Bielas amplified the adaptor-ligated fragments via PCR. EX1013 (Bielas) at 0026, 0031. As I discuss above, these PCR reactions, in addition to amplifying the fragments, resulted in removing the adaptor-ligated fragments from the circular vector. EX1013 (Bielas) at 0026–72, 0031.

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

507. Bielas discloses this limitation. As I explain above in Section VI.D.1, in Examples 1 and 3, Bielas sequenced the adaptor-ligated fragments on an Illumina NGS device. EX1013 (Bielas) at 0027, 0031. Sequencing on these Illumina platforms would have generated a population of sequencing reads.

5) "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;"

508. Bielas discloses this limitation. As I discuss in detail above in Section VI.D.1, Bielas identified duplicate sequencing reads using the sequence of the Cypher and the sequence of the DNA fragment in both Example 1 and Example 3. In Example 1, Bielas used the Cypher to "computationally deconvolute the sequencing data and map all the sequence reads to single molecules (i.e., distinguish PCR and sequencing errors from real mutations." EX1013 (Bielas) at 0027. Bielas also aligned the sequence of the DNA fragment with the then-known human genome. EX1013 (Bielas) at 0027. Using the Cypher sequence and sequence of the DNA fragment, Bielas identified duplicate sequencing reads and then collapsed those duplicates into "single molecules" in order to identify "tumor heterogeneity at the single-nucleotide level and reveal tumors have a mutator phenotype." EX1013 (Bielas) at 0027.

509. In Example 3, Bielas grouped sequence reads into families and created a consensus sequence first based on the Cypher. EX1013 (Bielas) at 0031. Then Bielas compared all of the sequences of the DNA fragments in the library in order to remove non-natural mutations from the consensus sequence so that true mutations could be identified. EX1013 (Bielas) at 0031–32. As I explain above in Section

VI.D.1.b of my declaration, Bielas diagrammed this process in Figure 3B, which I have reproduced below with my annotations. EX1013 (Bielas) at 0043, 0031–32.

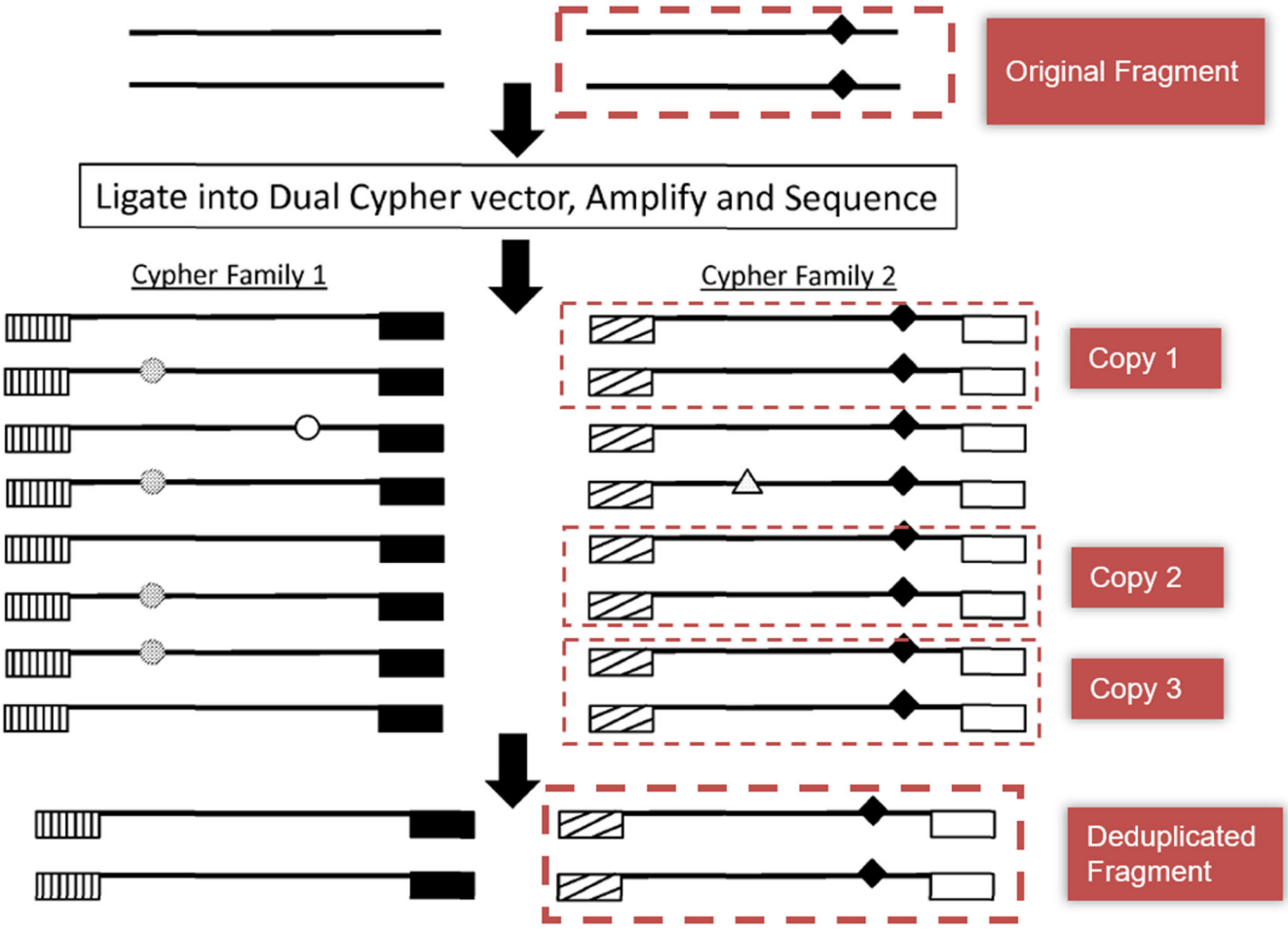


Fig. 3B

6) "wherein the indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides;"

510. Bielas discloses this limitation. As I explain in Section VI.D.1, Bielas disclosed the use of a sample index in both Example 1 and Example 3. Bielas further explained that these sample indexes are "a unique index sequence . . . specific for a

particular sample," and that "if ten different samples are being pooled in preparation for multiplex sequencing, then ten different index sequences will be used such that ten dual cypher vector libraries are used in which each library has a single, unique index sequence identifier." EX1013 (Bielas) at 0015–16. Bielas thus teaches that an "indexing site is unique amongst a subset of the plurality of nucleic acid fragments and is an index for multiple polynucleotides." EX1003 at 25:56–58.

7) "and wherein the sequence of the identifier site is variable in sequence content in a plurality of adaptors."

511. Bielas discloses this limitation. As I explain above in Section VI.D.1, Bielas teaches that Cyphers/identifier sites are comprised of random nucleotides. EX1013 (Bielas) at 0007–08. For instance, Bielas explained that a Cypher with the formula "5′-NNNNNNN-3′" would be comprised of 7 nucleotides "wherein N may be any naturally occurring nucleotide." EX1013 (Bielas) at 0016. In Example 3, Bielas provided the adaptor sequence, including a Cypher represented as "NNNNNNN," or 7 random nucleotides. EX1013 (Bielas) at 0028–29. Because Bielas teaches that each Cypher contains random nucleotides in each adaptor, the Cypher "is variable in sequence content in a plurality of adaptors." EX1003 at 25:58–60. Indeed, in Bielas's example of sample indexes, Bielas further states that "each library has a single, unique index sequence identifier (but each library has a plurality of random cyphers)." EX1013 (Bielas) at 0016.

b. To The Extent Bielas Does Not Anticipate Claim 1, Claim 1 Would Have Been Obvious Based On Bielas.

512. All of the elements of the method of claim 1 arranged as in the claim are taught in Bielas as I explain above. To the extent that Example 1 and Example 3 do not teach every element of claim 1 arranged as in the claim so as to anticipate claim 1, at a minimum, Bielas renders claim 1 obvious for the same reasons I discuss above in Sections VI.D.1 and VI.D.2.a.

513. To the extent that Tecan contends that Bielas Examples 1 and 3 do not anticipate the limitations relating to sample indexing and multiplexing, including a sample index and an indexing primer binding site, Bielas renders those limitations obvious. Bielas expressly contemplates the use of sample indexes and teaches that multiplexing is a feature of the Bielas methods. EX1013 (Bielas) at 0015–16, 0022, 0025. Moreover, Bielas used standard Illumina sequences that would have allowed the NGS device to read the sample index (either using the P5 graft sequence as a primer or another indexing primer). The POSA would have recognized that the use of the Illumina devices disclosed in Bielas in combination with the standard sequences Bielas used, meant that the methods could be easily used with standard Illumina sample indexes and adaptor sequences. Given these express teachings in Bielas itself, the POSA would have had a reasonable expectation of success in practicing the methods of Bielas with multiplexing.

514. To the extent that Tecan contends that Bielas Examples 1 and 3 do not expressly disclose an identifier site with a length of 1 to 8 nucleotides, Bielas renders that length of identifier site obvious. The range of 1 to 8 nucleotides would have been obvious to the POSA because Bielas used a Cypher of 7 nucleotides in Example 3. EX1013 (Bielas) at 0028.

515. In addition, as I discuss above in ¶¶ 265-266 in connection with my analysis of why Iafrate and Kivioja render an identifier site of 1 to 8 nucleotides obvious, at the priority date it would have been obvious to the POSA that the molecular tag could be designed to be a number of different nucleotides in length, and that it would have been beneficial to alter the length depending on certain experimental conditions. When the molecular tag sequence was read along with the fragment sequence, as in Bielas, that the POSA would have recognized that the length of the tag should be kept to a minimum so as to maximize the amount of the fragment sequence within the read length of the NGS device. This was particularly true when sequencing on an Illumina device in 2013, given the limited read length of the most common Illumina models. Accordingly, the POSA would have wanted to keep the molecular tag length as short as possible while ensuring a sufficient number of molecular tags to make each fragment unique.

516. The POSA would have recognized, as taught in McCloskey and other prior art references, which I discuss in Section IV.K.2.b and ¶¶ 264-266 of my

declaration, that the number of unique molecular tags was a simple function of the length of the molecular tag (i.e. 4^n). EX1022 (McCloskey) at 0004; EX1026 (Casbon) at 0001, 0005. In this regard, as I discuss in ¶¶ 264-266 above, I note that the length of the molecular tag is a "result-effective variable" (as I understand that term is defined in the law set out in Section III.D. above)—where the length of the tag has a predictable relationship with the total number of unique tags, which in turn has a predictable relationship with the expected number of unique fragments of interest in a particular sample. The POSA thus would have been motivated the decrease the length of the molecular tag if significantly fewer than 4^n unique fragments were expected.

517. This is consistent with the teachings of Bielas, which explains that the length of the Cypher can vary depending on the application, and that the Cypher can "comprise a length ranging from about 5 nucleotides to about 50 nucleotides (or about 5 nucleotides to about 10 nucleotides, or a length of about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, or about 20 nucleotides." EX1013 (Bielas) at 0005. Bielas discloses other ranges of possible cypher lengths at page 0014, and there again expressly discloses cyphers of about 6, about 7, or about 8 nucleotides. At page 0023, Bielas discloses s construct very similar to that shown in Figure 2 and states that a range of 7 to 9 is preferred. Thus, Bielas teaches that the Cypher/identifier

site can be 6, 7, or 8 nucleotides, rendering the range of 1 to 8 obvious to the POSA. Moreover, the ranges disclosed in Bielas at pages 0005, 0010, and 0014, including 5 to 50, 5 to 10, 5 to 12, 6 to 10, 7 to 12, 7 to 9, and 6 to 8 all overlap with the claimed range of 1 to 8 nucleotides and show that the claimed range would have been obvious to the POSA.

518. Bielas goes on to explain that one should select the appropriate number of nucleotides for a particular application based on exactly the known relationship that I explain above, and the authors note that the number of cyphers can be less than the total number of fragments "provided that each such target nucleic acid molecule does not have the same double-stranded cypher sequence as any other such target nucleic acid molecule." EX1013 (Bielas) at 0005. Based on this teaching in Bielas and the other prior art with the same teachings, the POSA would understand that the length of the Cypher could be routinely optimized to a particular application and the claimed range of 1 to 8 nucleotides would have been obvious.

519. As I explain above in connection with my opinion that this limitation would have been obvious over Iafrate and Kivioja, the '399 patent does not disclose any special or unexpected properties for an identifier site of 1 to 8 nucleotides. To the contrary, the '399 patent does not disclose any data whatsoever from the use of the claimed methods. Thus, I do not understand Tecan to have provided any evidence that the claimed range of 1 to 8 nucleotides has any special or desirable

properties, or any evidence that the claimed range of 1 to 8 nucleotides performs any differently than the specific lengths (including 6, 7, and 8 nucleotides) or the various ranges disclosed in Bielas.

c. Claim 2: "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."

520. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

521. Bielas further discloses the additional limitations of claim 2. As I explain above in Section VI.D.1.a, in Example 1, Bielas used the Cypher sequence to "deconvolute" sequencing data into "single molecules" and then mapped the sequences of the DNA fragments to the then-known human genome in order to identify mutations in the original DNA fragments. EX1013 (Bielas) at 0027. That is a process of identifying and removing duplicate sequencing reads. When reads are collapsed into "single molecules," the other reads corresponding to those other molecules are being removed from the population of sequencing reads. EX1013 (Bielas) at 0027.

522. Bielas also removed duplicate sequencing reads in Example 3 as I discuss above in Section VI.D.1.b. In that example, 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 then used the sequence of the DNA fragments to identify and remove (or "computationally eliminate," in Bielas's terms) "errors introduced during library preparation" and errors introduced "during sequencing." EX1013 (Bielas) at 0031–32. After the consensus sequence is generated, the individual reads that were used to generate the consensus sequence are removed from the population of sequencing reads.

523. To the extent that Bielas does not anticipate this claim, Bielas renders it obvious. The POSA would have understood that one of the purposes of the method described in Bielas was to streamline and simplify further data analysis. To accomplish this, the POSA would have understood that it was desirable to remove duplicate sequencing reads from the population of sequencing reads prior to further analysis, and would have had a reasonable expectation of success in doing so as part of NGS data analysis. For these reasons, the POSA would have had a reasonable expectation of success in practicing the methods of claim 2.

d. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the indexing site."

524. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

525. The ordinary meaning to the POSA in this context of sequencing the

identifier site with the indexing site is that both sequences are sequenced in a single SBS read on the NGS device. While Bielas sequenced the identifier site with the nucleic acid fragment, and sequenced the indexing site separately, the POSA would have recognized that whether the indexing site and the identifier site were sequenced together in one SBS read or separately was simply a matter of the design of the adaptor. It would have been obvious to the POSA that the indexing sequence could have been placed on the 3′ side of the sequencing primer binding site in the adaptor so as to sequence the indexing site along with the identifier site and the fragment. The order of the various elements in the adaptor would simply have been a design choice for the POSA, and the POSA would have had a reasonable expectation of success in modifying the adaptor as I discuss in this paragraph. The POSA would have had a reasonable expectation of success of organizing the features of an adaptor such that the sample index/indexing site and molecular tag/identifier site were sequenced together.

526. Moreover, I do not understand the ’399 patent to disclose anything unexpected or beneficial about the placement of the indexing site and the identifier site in the adaptor so as to sequence them in one SBS read or two. To the contrary, the POSA would have recognized that useful sequencing data could be obtained using adaptors in either arrangement.

e. Claim 5: "The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample."

527. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

528. Bielas further anticipates the additional limitations of claim 5. As I discuss above in Sections VI.D.1.a-b, Bielas used sample indexes in both Example 1 and Example 3. The purpose of these sample indexes was to permit the pooling of nucleic acid fragments from more than one sample. Once libraries are uniquely labeled with a sample index, Bielas teaches they can be pooled together for sequencing. EX1013 (Bielas) at 0015–16. Accordingly, the POSA would have recognized that these examples taught the additional limitation that the plurality of nucleic acid fragments is generated from more than one sample.

529. To the extent that Bielas does not anticipate sample pooling and multiplex sequencing, Bielas renders this limitation obvious by expressly teaching the methods are compatible with sample indexes and multiplex sequencing. EX1013 (Bielas) at 0015–16. Based on these teachings in Bielas, the POSA would have had a reasonable expectation of success that the methods in Bielas could be used with sample pooling and multiplexing.

f. Claim 6: "The method of claim 5, wherein the nucleic acid fragments from each sample have the same indexing site, and each sample has a unique indexing site."

530. The limitations of claims 1 and 5 are anticipated and/or obvious based on Bielas for the reasons I describe above.

531. Bielas further anticipates the additional limitations of claim 6. As I discuss above in connection with claims 1 and 5, Bielas discloses the use of a sample index in both Example 1 and Example 3. EX1013 (Bielas) at 0027–29. Bielas further explains that sample indexes are used so that fragments from each particular sample have "a single, unique index sequence identifier." EX1013 (Bielas) at 0015–16. As Bielas explains by way of example, "if ten different samples are being pooled in preparation for multiplex sequencing, then ten different index sequences will be used such that ten dual cypher vector libraries are used in which each library has a single unique index sequence identifier (but each library has a plurality of random cyphers)." EX1013 (Bielas) at 0015–16.

532. To the extent that Bielas does not anticipate the additional limitations of claim 6, Bielas renders those limitations obvious based on Bielas's express teachings of sample indexing with a unique sample index. EX1013 (Bielas) at 0015–16. Based on these teachings in Bielas, the POSA would have had a reasonable expectation of success that the methods in Bielas could be used with sample pooling

and multiplex sequencing.

g. Claim 7: "The method of claim 6, wherein the sequencing reads are separated based on the indexing site."

533. The limitations of claims 1, 5, and 6 are anticipated and/or obvious based on Bielas for the reasons I described above.

534. Bielas further discloses the additional limitations of claim 7. As explained above with respect to claims 1, 5, and 6, in Example 1 and Example 3, Bielas teaches using a sample index incorporated in an adaptor. Bielas sequenced the adaptor-ligated DNA fragments in both examples on Illumina devices. The POSA would have known that the data output from sequencing on those Illumina devices is a FASTQ file with the pooled samples demultiplexed, i.e. separated based on indexing site. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35.

535. To the extent that Bielas does not anticipate the additional limitations of claim 7, Bielas renders those limitations obvious. It would have been obvious to the POSA to demultiplex data and separate reads by sample, and as I discuss above, this was standard in the prior art use of Illumina devices. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35. The POSA would have had a reasonable expectation of success in demultiplexing reads as part of the methods described in Bielas.

h. Claim 8: "The method of claim 7, wherein the separation of sequencing reads is performed prior to step d)."

536. The limitations of claims 1, 5, 6, and 7 are anticipated and/or obvious based on Bielas for the reasons I describe above.

537. Bielas further discloses the additional limitations of claim 8. As explained above in connection with claims 1, 5, 6, and 7, in Example 1 and Example 3, Bielas teaches using a sample index incorporated in an adaptor and sequenced the adaptor-ligated fragments on an Illumina device. The POSA further would have understood that these Illumina devices would have output the data as demultiplexed FASTQ files. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35. As I discussed above in Section ¶ 97, this demultiplexing step would occur prior to further data analysis, EX1062 (Smith) at 0034–38; EX1063 (Won) at 0003–04, including prior to deduplicating the samples using the Cypher and fragment sequence, EX1054 (Loeb) at 0012–15.

538. To the extent that Bielas does not anticipate the additional limitations of claim 8, Bielas renders those limitations obvious. It would have been obvious to the POSA to demultiplex data and separate reads by sample prior to further data analysis, and as I discuss above, this was standard in the prior art use of Illumina devices. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–38; EX1063 (Won) at 0003-04; EX1054 (Loeb) at 0012–15. The POSA would have had a

reasonable expectation of success in demultiplexing reads prior to further analysis as part of the methods described in Bielas.

539. In addition, I do not understand the '399 patent to disclose any particular benefit to demultiplexing the data prior to de-duplicating it as required by claim 8. To the contrary, the POSA would have recognized that these respective data analysis operations could be performed in either order depending on the particular application or experimental design.

i. Claim 9: "The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments."

540. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

541. Bielas discloses the additional limitations of claim 9. Bielas's proof of concept experiments recited in Example 1 and Example 3 both used DNA fragments. EX1013 (Bielas) at 0026–32. Accordingly, Bielas anticipates claim 9, or, to the extent that Bielas only renders the limitations of claim 1 obvious instead of anticipated, renders claim 9 obvious. Based on the methods in Bielas, the POSA would have had a reasonable expectation of success of practicing the method with the nucleic acids specified in claim 9.

j. Claim 10: "The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."

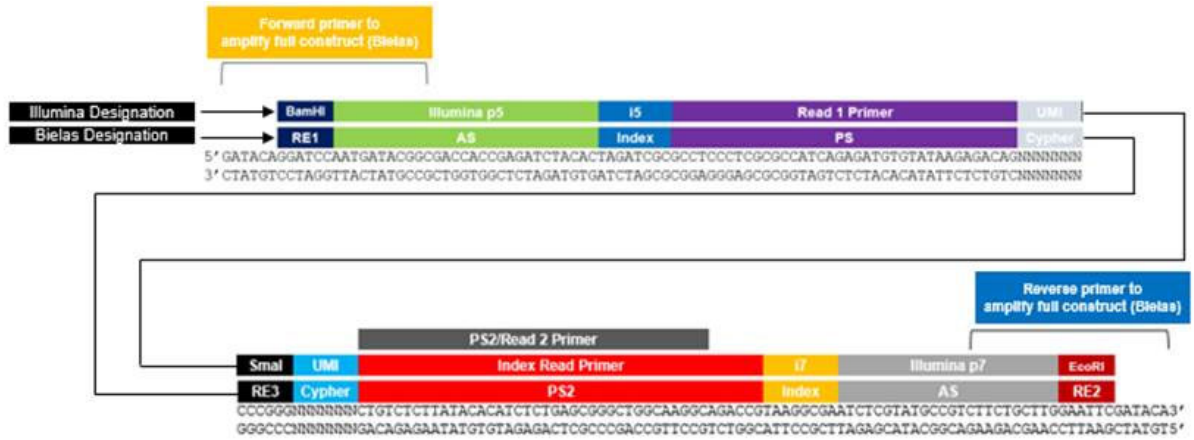
542. The limitations of claims 1 and 9 are anticipated and/or obvious based on Bielas for the reasons I describe above.

543. Bielas discloses the additional limitations of claim 10. Bielas's proof of concept experiments recited in Example 1 and Example 3 both used genomic DNA fragments. EX1013 (Bielas) at 0026–32. Accordingly, Bielas anticipates claim 10, or, to the extent that Bielas only renders the limitations of claim 1 obvious instead of anticipated, renders claim 10 obvious. Based on the methods in Bielas, the POSA would have had a reasonable expectation of success of practicing the method with the nucleic acids specified in claim 10.

k. Claim 11: "The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length."

544. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

545. Bielas discloses the additional limitations of claim 11. As explained, in Example 3, Bielas disclosed the sequence of the adaptor used, which included an 8-nucleotide sample index, as annotated in the figure below. EX1013 (Bielas) at 0028–29. This disclosure falls within the claimed range of 2 to 8 nucleotides and thus anticipates it.



546. Furthermore, Bielas disclosed the use of standard Illumina adaptor sequences. EX1013 (Bielas) at 0028–29. The POSA would also have known that, by 2013, Illumina's adaptors for multiplexed sequencing included a standard 8-nucleotide sample index, the so-called i5 index. EX1056 (Illumina Letter) at 0003; EX1057 (Nextera Prep Guide) at 0044.

547. To the extent that Bielas does not anticipate this limitation, it renders a sample index of 2 to 8 nucleotides obvious, including because Bielas used an 8 nucleotide sample index in Example 3 and disclosed the use of standard Illumina sample indexes.

548. In addition, Bielas discloses that sample indexes may comprise "a length ranging from about 4 nucleotides to about 25 nucleotides." EX1013 (Bielas) at 0015. This disclosure of a range of sample index nucleotides in Bielas overlaps the claimed range and thus renders it obvious. I do not understand the '399 patent to provide any evidence of unexpected properties for the claimed range of sample

index lengths.

549. Furthermore, even if Bielas did not teach indexing sites within the range of 2 to 8 nucleotides, it would have been obvious to the POSA to arrive at a sample index in that range via routine optimization. At the priority date, the POSA would have understood that sample indexes could be designed to different lengths depending on the design needs of a particular application. The POSA would have recognized that the length of the sample index was simply dependent on the number of samples and the length of sequence required to ensure to the POSA that samples would remain uniquely labeled even accounting for PCR and sequencing error. To multiplex more samples, the POSA could use longer sample indexes, and with fewer samples, the POSA could use shorter sample indexes. Based on this predictable relationship, the POSA would have arrived at a sample index between 2 and 8 nucleotides based on routine optimization.

550. In addition, I do not understand the '399 patent to teach any unexpected results or benefits of the particular length of 2 to 8 nucleotides for an indexing site. To the contrary, this range of nucleotides is simply arbitrary and the POSA would not have recognized any special benefit to that length over, for example, a sample index of 9 nucleotides.

l. Claim 14: "The method of claim 1, wherein the identifier site is 8 nucleotides in length."

551. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

552. Bielas discloses the additional limitations of claim 14 and a Cypher length of 8 nucleotides would have been obvious to the POSA based on Bielas. At multiple points, Bielas states that in some embodiments, the Cypher may be 8 nucleotides in length. EX1013 (Bielas) at 0014, 0015. At page 0014, Bielas explains that "[i]n certain embodiments a cypher preferably has a length of about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, or about 20 nucleotides." At page 0015, Bielas states that "[i]n certain embodiments, both the X^a and X^b cypher each comprise 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, 11 nucleotides, 12 nucleotides, 13 nucleotides, 14 nucleotides, 15 nucleotides, 16 nucleotides, 17 nucleotides, 18 nucleotides, 19 nucleotides, or 20 nucleotides." Bielas goes on to explain on page 0015 that "[i]n certain other embodiments," "the X^a cypher [i.e. the 5′ cypher] comprises 8 nucleotides."

553. Bielas is clear that the length of the Cypher length can be routinely optimized for different applications. As I discuss above in connection with claim 1 in Section VI.D.2.a.2.c, Bielas explains that the length of the cypher can be routinely optimized based on the resulting number of cyphers (defined by the formula 4^n) and the expected number of unique fragments. As I discuss further above in ¶ 263, at

the priority date, the POSA would have been motivated to keep the length of the identifier site short in order to maximize the length of the fragment sequence read, particularly given the limited read lengths of the Illumina devices used by Bielas.

554. As I further discuss in ¶¶ 264-266 above, the length of the identifier site is a "result-effective variable" (as I understand that term is defined in the law set out in Section III.D above) wherein there was a known relationship between the variable (length of the identifier site) and the parameter being optimized (the number of labels necessary to uniquely tag the number of unique original fragments expected). Based on this routine optimization, the POSA would have considered an identifier site of 8 nucleotides obvious. Based on the above, the POSA would have had a reasonable expectation of success of practicing the method with an 8-nucleotide identifier site.

555. I further note that the '399 patent does not teach that there was anything unexpectedly desirable or critical about an identifier site of specifically 8 nucleotides. To the contrary, the '399 patent does not provide any data whatsoever from experiments using the claimed methods. The POSA would not expect any such data—there is nothing particular special or valuable about an identifier site of 8 nucleotides (*i.e.* 4^8, or 65,536 unique labels). That number of labels would be appropriate for some applications (with fewer unique fragments) but not appropriate for others (with more unique fragments), and the POSA would recognize that the determination of the appropriate length of the identifier site would be a matter of

routine optimization.

m. Claim 15: "The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site."

556. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

557. Bielas discloses the additional limitations of claim 15. As explained above, the indexing primer binding site disclosed in Example 1 and Example 3 are also the sequencing primer binding site ("PS1") (for sequencing on the Genome Analyzer II) and the Illumina flow cell P5 graft sequence ("AS") (for sequencing on the Illumina MiSeq), respectively. The PS1 and AS sequences are standard Illumina sequences that are the same in every adaptor. The indexing primer binding sites disclosed in Example 1 and Example 3 are thus universal indexing primer binding sites. Accordingly, Bielas anticipates claim 15, or, to the extent that Bielas only renders the limitations of claim 1 obvious instead of anticipated, renders claim 15 obvious. Based on the methods in Bielas, the POSA would have had a reasonable expectation of success of practicing the method of claim 15.

n. Claim 16: "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site."

558. The limitations of claim 1 are anticipated and/or obvious based on Bielas for the reasons I describe above.

559. Bielas discloses the additional limitations of claim 16. As explained above, the target sequence primer binding sites disclosed in Example 1 and Example 3 are standard Illumina sequences for sequencing on the Illumina Genome Analyzer II and the Illumina MiSeq, respectively. These sequences are the same in each adaptor ligated to the DNA fragments in Example 1 and Example 3. The target sequence primer binding sites disclosed in Example 1 and Example 3 are thus universal target sequence primer binding sites. Accordingly, Bielas anticipates claim 16, or, to the extent that Bielas only renders the limitations of claim 1 obvious instead of anticipated, renders claim 16 obvious. Based on the methods in Bielas, the POSA would have had a reasonable expectation of success of practicing the method of claim 16.

o. Claim 17: "The method of claim 6, wherein the method further comprises separating fragments or sequencing reads based on indexing site."

560. The limitations of claims 1, 5, and 6 are anticipated and/or obvious based on Bielas for the reasons I described above.

561. Bielas discloses the additional limitations of claim 17. As explained, in Example 1 and Example 3, Bielas teaches using a sample index in an adaptor. Bielas teaches that the adaptor-ligated DNA fragments in Example 1 and Example 3 were sequenced on an Illumina Genome Analyzer II and an Illumina MiSeq, respectively. EX1013 (Bielas) at 0027, 0031. As I discussed in ¶ 97 and Section

VI.D.2.h above, the POSA would have known that the data output from sequencing on those Illumina devices is a FASTQ file with the pooled samples demultiplexed, i.e. with the fragments or sequencing reads separated based on indexing site. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034–35. For these reasons, Bielas anticipates claim 17.

562. To the extent that Bielas does not anticipate claim 17, the claim would have been obvious based on Bielas. It would have been obvious to the POSA to separate reads based on sample index prior to further data analysis, consistent with standard Illumina workflows. EX1060 (Meyer & Kircher) at 0009; EX1062 (Smith) at 0034-35. For these reasons, the POSA would have had a reasonable expectation of success of practicing the method of claim 17.

3. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The ’357 Patent.

563. Bielas teaches every element of the method claimed in claims 1–4, 6–7, and 9–10 of the ’357 patent and thus anticipates the claims and/or renders the claims obvious. As I discuss above in connection with Iafrate and Kivioja in Section VI.C.7 of my declaration, the ’357 claims are in most regards straightforwardly broader in scope than the ’399 claims. Thus, my arguments with respect to why Bielas anticipates the claims of the ’399 patent apply equally to the broader claims of the ’357 patent.

564. Specifically, Bielas teaches a method using adaptors containing all of the features recited in the claims of the ’357 patent, including a sample index, an identifier site (which Bielas refers to as a “Cypher”), and a primer binding site. EX1013 (Bielas) at 0041, 0038. Bielas further teaches using the Cypher sequence and sequence of the nucleic acid fragment to identify PCR copies of original fragments after an NGS read, which necessarily requires identifying a duplicate sequencing read. EX1013 (Bielas) at 0026-32.

a. Bielas Anticipates Claim 1.

565. For the reasons stated in Section VI.D.1 of my declaration and the following reasons, Bielas anticipates claim 1 of the ’357 patent.

1) “A method for detecting duplicate sequencing reads, the method comprising:”

566. The preamble to claim 1 of the ’357 patent and the preamble to claim 1 of the ’399 patent are directed at the same method. The POSA would have understood that there was no substantive difference between the two preambles, even though the ’357 preamble is plural (i.e. “detecting duplicate sequencing reads”) and the ’399 preamble is singular (i.e. “detecting a duplicate sequencing read”).

’357 Preamble	’399 Preamble
1. “A method for detecting duplicate sequencing reads” EX1004 at 25:36.	1. “A method for detecting a duplicate sequencing read from a population of sample sequencing reads” EX1003 at 25:37-38.

567. Because the ’357 preamble is substantively identical to the ’399 preamble, Bielas discloses this limitation for the reasons stated in Section VI.D.2.a.1 of my declaration.

2) “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”

568. The first step in claim 1 of the ’357 patent is broader than the first step in claim 1 of the ’399 patent, as shown in the table below.

’357 Ligation Step	’399 Ligation Step
“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” EX1004 at 25:38–42.	“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 each 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” EX1003 at 25:39–46.

569. Each limitation in the ligation step in the ’357 patent, including the features of the adaptor, is also present in the ’399 patent. The difference between the two is that the ’357 limitations are broader. First, while the ’399 patent requires ligation “to a 5′ end of each nucleic acid fragment,” EX1003 at 25:39, the ’357 patent simply requires ligation to nucleic acid fragments, whether on the 5′ or 3′ end,

EX1004 at 38-39. Second, while the '399 patent requires that adaptors feature specific primer binding sites, "an indexing primer binding site" and "a target sequence primer binding site," EX1003 at 25:42, 46, the '357 patent requires simply that adaptors feature "a primer binding site," EX1004 at 25:42. Third, while the '399 patent requires the identifier site to be "between 1 and 8 nucleotides," EX1003 at 25:44, the '357 patent requires the identifier site to be "about 1 to about 8 nucleotides," EX1004 at 25:40.

570. I do not understand the '357 patent's characterization of an identifier site as "a unique identifier," EX1004 at 25:39, to be different than the "identifier site" in the '399 patent, EX1003 at 25:44. The '357 patent does not define "unique identifier;" rather, it defines "identifier site," EX1004 at 4:49–52, in the same way the '399 does, EX1003 at 4:43–46. And certain dependent claims in the '357 patent use "identifier site" rather than "unique identifier." EX1004 at 25:50, 26:35.

571. Rather, I understand the qualifier "unique" to mean that "the sequence of the identifier site is variable in sequence content" in each adaptor, as is required in claim 1 of the '399 patent. EX1003 ('399 patent) at 25:59–61.

572. Because each limitation in the ligation step of the '357 patent is also present in the '399 patent, Bielas discloses this limitation for the reasons stated in Sections VI.D.2.a.2 and VI.D.2.a.7 of my declaration.

3) "amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences"

573. The second and third steps in claim 1 of the '357 patent overlap with the second and third steps in claim 1 of the '399 patent, as shown in the table below.

'357 Amplification and Sequencing Step	**'399 Amplification and Sequencing Step**
"amplifying the adaptor-ligated fragments into amplicons; sequencing the amplicons to produce sequence reads that include identifier and target sequences" EX1004 at 25:43–45.	"b) amplifying the adapter-nucleic acid fragment ligated products; c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products" EX1003 at 25:47–51.

574. Despite slight differences in the language, there is no substantive difference between these steps of the '357 and '399 patents. The '399 patent does not use the term "amplicons," but the POSA would have understood "amplicons" to be the amplified adaptor-ligated nucleic acid fragments. (The '357 patent does not define otherwise define "amplicons" differently from its ordinary meaning.) The POSA would also have understood "sequencing the amplicons to produce sequence reads," EX1004 at 25:44, to be equivalent to "generating a population of sequencing reads," EX1003 at 25:49, i.e. both the '357 and the '399 patents require sequencing after amplification.

575. Finally, while the '399 patent does not expressly say that sequencing reads "include identifier and target sequences," as the '357 patent does, EX1004 at

25:46–47, the ’399 patent necessarily requires it. Otherwise remaining steps of the ’399 patent would not have worked; for example, without the identifier and target sequences, duplicate reads could not be identified as “comprising the same identifier site and nucleic acid fragment.” EX1003 at 25:52–55.

576. Because each limitation in the amplification and sequencing steps of the ’357 patent are also present in the ’399 patent, Bielas discloses these limitations for the reasons stated in Sections VI.D.2.a.3–4 of my declaration.

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

577. The fourth step in claim 1 of the ’357 patent overlaps with the fourth step in claim 1 of the ’399 patent, as shown in the table below.

’357 Identification Step	’399 Identification Step
“identifying sequence reads with identical identifier and target sequences as duplicates” EX1004 at 25:46–47.	“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” EX1003 at 25:52–55.

578. Despite slight differences in the language, both the ’357 patent and the ’399 patent require the same thing: identify duplicate sequencing reads as having identical (or the same) identifier site and nucleic acid fragment sequences. The POSA would have understood “identical,” as used in the ’357 patent, EX1004 at 25:46, to be equivalent to “same,” as used in the ’399 patent, EX1003 at 25:43. In

other words, the POSA would have understood that "identical identifier and target sequences," EX1004 at 25:46–47, is the same thing as "the same identifier site and nucleic acid fragment," EX1003 at 25:53–54. The '357 patent even describes duplicate sequencing reads in the language of the '399 patent: "A duplicate sequencing read can be a sequencing read with the same identifier site and target sequence as another sequencing read in the population of sequencing reads." EX1004 at 4:12–15.

579. Because the identification step of the '357 patent is also present in the '399 patent, Bielas discloses this limitation for the reasons stated in Section VI.D.2.a.5 of my declaration.

b. To The Extent Bielas Does Not Anticipate Claim 1, Claim 1 Would Have Been Obvious Based On Bielas.

580. All of the elements of the method of claim 1 arranged as in the claim are taught in Bielas as I explain above. To the extent Example 1 and Example 3 do not teach every element of claim 1 arranged as in the claim so as to anticipate claim 1, at a minimum, Bielas renders claim 1 obvious. As I explained above in Section VI.D.3.a of my declaration, claim 1 of the '357 patent entirely overlaps with, and is broader than, claim 1 of the '399 patent. Thus, to the extent Bielas does not anticipate claim 1 of the '357 patent, the claim is obvious for the same reasons claim 1 of the '399 patent is obvious, as I discussed above in Section VI.D.2.b.

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

581. The limitations of claims 1 and 2 are anticipated and/or obvious based on Bielas for the reasons I describe above.

582. Claim 2 of the '357 patent claims the same limitation as claim 2 of the '399 patent, as shown in the table below.

'357 Claim 2	'399 Claim 2
2. "The method of claim 1, further comprising removing a duplicate read from the sequence reads." EX1004 at 25:48–49.	2. "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." EX1003 at 25:61–64.

583. Because the limitation in claim 2 of the '357 patent claim is equivalent to the limitation in claim 2 of the '399 patent, Bielas anticipates claim 2 for the reasons stated in Section VI.D.2.c of my declaration. To the extent Bielas does not anticipate this limitation, Bielas renders it obvious and the POSA would have had a reasonable expectation of success for the reasons stated in Section VI.D.2.c of my declaration.

d. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence."

584. The limitations of claims 1 and 3 are anticipated and/or obvious based on Bielas for the reasons I describe above.

585. As I explained in relation to claim 3 of the ’399 patent, the ordinary meaning to the POSA of sequencing the identifier site with the target sequence is that both sequences are sequenced in a single SBS read on the NGS device. Under that interpretation, Bielas anticipates claim 3. As I discuss above in Section VI.D.1, in both Example 1 and Example 3, Bielas disclosed the use of adaptors with the molecular tag and nucleic acid fragment to the 3′ side of the sequencing primer binding site. The placement of the sequencing primer binding site relative to the molecular tag and nucleic acid fragment means the molecular tag and nucleic acid fragment would have been sequenced in a single SBS read.

586. To the extent this claim is not anticipated by Bielas, it would have been obvious. As I explained in Section VI.D.1.b of my declaration, Bielas disclosed the adaptor sequence used in Example 3. In that adaptor, the Cypher/identifier site is to the 3′ side of the target sequence primer binding site, meaning the identifier site and target sequence were read together. Furthermore, as I explained in Section VI.D.1.a of my declaration, the adaptor disclosed in Figure 2 in Bielas, and used in Example 1, has the Cypher to the 3′ side of the target sequence primer binding site (“PS1” in Figure 2). As in Example 3, the identifier site and target sequence were read together in Example 1. Thus, claim 3 would have been obvious based on Bielas. For these reasons, the POSA would have had a reasonable expectation of success in practicing the method of claim 3.

587. Claim 3 would further have been obvious because the POSA would have recognized that whether the identifier site and target sequence were sequenced together in one SBS read or separately was simply a matter of the design of the adaptor. It would have been obvious to the POSA that the identifier site could have been placed within the adaptor such that the identifier site and nucleic acid fragment were sequenced together. The order of the various elements in the adaptor would simply have been a design choice for the POSA, and the POSA would have had a reasonable expectation of success in modifying the adaptor as I discuss above.

588. Moreover, I do not understand the '357 patent to disclose anything unexpected or beneficial about the placement of the identifier site within the adaptor so as to sequence the identifier site and target sequence in one SBS read or two. To the contrary, the POSA would have recognized that useful sequencing data could be obtained whether the identifier site was sequence with or without the target sequence.

e. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence."

589. The limitations of claims 1 and 4 are anticipated and/or obvious based on Bielas for the reasons I describe above.

590. As I discuss above in Section VI.D.1, in both Example 1 and Example 3, Bielas disclosed the use of adaptors on either end of the nucleic acid fragment.

The adaptors on either side of the nucleic acid fragments contained standard Illumina sequences allowing the adaptor-nucleic acid fragment to bind to the Illumina flow cells in the Illumina devices used in both examples. EX1013 (Bielas) at 0027–31, 0041. The POSA would have known that the use of Illumina devices with adaptors on either end allowed Bielas to use paired-end sequencing, as I describe above in Section IV.K.3. With paired-end sequencing Bielas sequenced the molecular tag in one adaptor (e.g., the 5′ adaptor) separately from the nucleic acid fragment sequencing read primed from the sequencing primer binding site in the opposite adaptor (e.g., the 3′ adaptor). Thus, Bielas disclosed sequencing the identifier site separately from the target sequence.

591. To the extent claim 4 is not anticipated by Bielas, it would have been obvious. The POSA would have understood Bielas's disclosure of standard Illumina sequences that allow the adaptor-nucleic acid fragments to anneal to the flow cell at both the 5′ and 3′ ends to teach the use of paired-end sequencing. The POSA would have further understood that with paired-end sequencing the molecular tag could have been sequenced separately from the nucleic acid fragment. For these reasons, the POSA would have had a reasonable expectation of success in practicing the methods of claim 4.

592. To the extent Claim 4 is not anticipated, at a minimum it would have been obvious based on Bielas for the reasons I explain in the prior section of my

declaration, Section VI.D.3.d. The order of the various elements in the adaptor would simply have been a design choice for the POSA, and the POSA would have had a reasonable expectation of success in modifying the adaptor to sequence the molecular tag separately from the nucleic acid fragment.

f. Claim 6: "The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments."

593. The limitations of claims 1 and 6 are anticipated and/or obvious based on Bielas for the reasons I describe above.

594. Claim 6 of the '357 patent claims the same limitation as claim 10 of the '399 patent, as shown in the table below.

'357 Claim 6	'399 Claim 10
6. The method of claim 1, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1004 at 26:39–40.	10. The method of claim 9, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments. EX1003 at 26:53–54.

595. In the '399 patent, claim 10 depends from claim 9, which limits the nucleic acid fragments to "DNA fragments, RNA fragments, or DNA/RNA fragments." EX1003 at 26:33–34. In other words, claim 10 of the '399 patent limits claim 9 to more specific types of DNA fragments. Claim 6 of the '357 patent provides for the same specificity but it depends from claim 1. I understand that these differences do not alter the substance of what is claimed, and thus I understand claim 6 of the '357 patent to claim the same limitation as claim 10 of the '399 patent.

596. Because the limitation in claim 6 of the ’357 patent is equivalent to the limitation in claim 10 of the ’399 patent, Bielas discloses this limitation for the reasons stated in Section VI.D.2.j of my declaration. To the extent that Bielas does not anticipate claim 6, Bielas renders it obvious and the POSA would have had a reasonable expectation of success for the reasons stated in Section VI.D.2.j of my declaration.

g. Claim 7: “The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site.”

597. The limitations of claims 1 and 7 are anticipated and/or obvious based on Bielas for the reasons I describe above.

598. Claim 7 of the ’357 patent overlaps with the limitation in claim 16 of the ’399 patent, as shown in the table below.

’357 Claim 7	’399 Claim 16
7. “The method of claim 1 wherein the primer binding site is a universal target sequencing primer binding site.” EX1004 at 26:41–42.	16. “The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.” EX1003 at 26:65–67.

599. Both claim 7 of the ’357 patent and claim 16 of the ’399 patent require “a universal target sequence primer binding site.” EX1004 at 26:41–42; EX1003 at 26:65–67. That claim 7 of the ’357 patent limits a “primer binding site,” as opposed to “the target sequence primer binding site,” as in the ’399 patent, does not alter my opinion because both ultimately specify that a primer binding site in the adaptor must

be "a universal target sequence primer binding site."

600. Thus, because the limitation in claim 7 of the '357 patent is the same as the limitation in claim 16 of the '399 patent, Bielas discloses this limitation for the reasons stated in Section VI.D.2.n of my declaration. To the extent that Bielas does not anticipate claim 7, Bielas renders it obvious and the POSA would have had a reasonable expectation of success for the reasons stated in Section VI.D.2.n of my declaration.

h. Claim 9: "The method of claim 1, wherein each unique identifier is 8 nucleotides in length."

601. The limitations of claims 1 and 9 are anticipated and/or obvious based on Bielas for the reasons I describe above.

602. Claim 9 of the '357 patent claims the same limitation as claim 14 of the '399 patent, as shown in the table below.

'357 Claim 9	'399 Claim 14
9. The method of claim 1, wherein each unique identifier is 8 nucleotides in length. EX1004 at 26:45–46.	14. The method of claim 1, wherein the identifier site is 8 nucleotides in length. EX1003 at 26:61–62.

603. Because the limitation in claim 9 of the '357 patent is equivalent to the limitation in claim 14 of the '399 patent, Bielas discloses this limitation for the reasons stated in Section VI.D.2.l of my declaration. To the extent that Bielas does not anticipate claim 9, Bielas renders it obvious and the POSA would have had a reasonable expectation of success for the reasons stated in Section VI.D.2.l of my

declaration.

i. Claim 10: "The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic acid fragment."

604. The limitations of claims 1 and 10 are anticipated and/or obvious based on Bielas for the reasons I describe above.

605. The '357 patent does not define "junction" but the ordinary meaning in this context is the location where the two sequences are joined. The '357 patent uses the term "junction" in this way; for example, in Figure 4 where the identifier site in the adaptor is joined to the nucleic acid fragment at what is described as "the insert junction." EX1004 ('357 patent) at 0010. Thus, I understand claim 10 of the '357 patent to require the identifier site to be located in the adaptor such that the identifier site is joined to the nucleic acid fragment.

606. Bielas discloses this limitation. As explained above in Section VI.D.1 of my declaration, Bielas constructed adaptors like those shown in Figure 2, reproduced below from EX1013 (Bielas) at 0041:

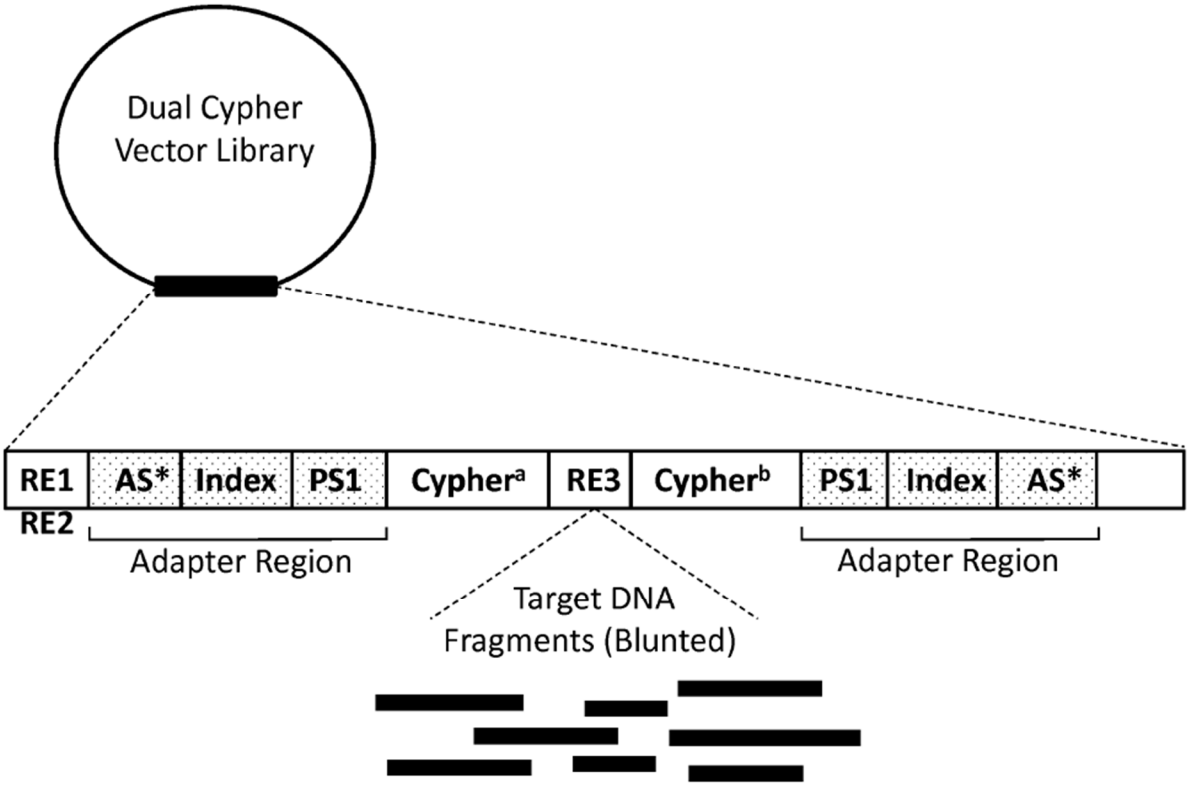
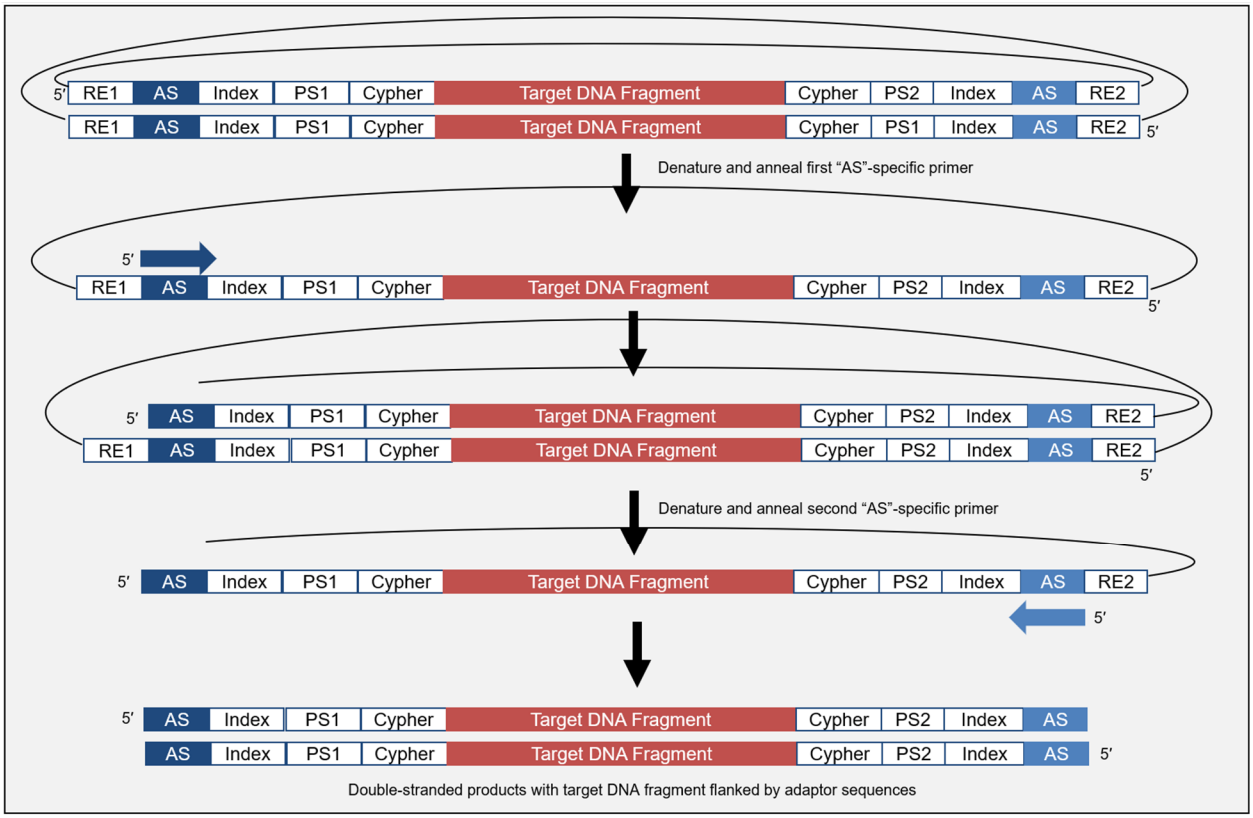
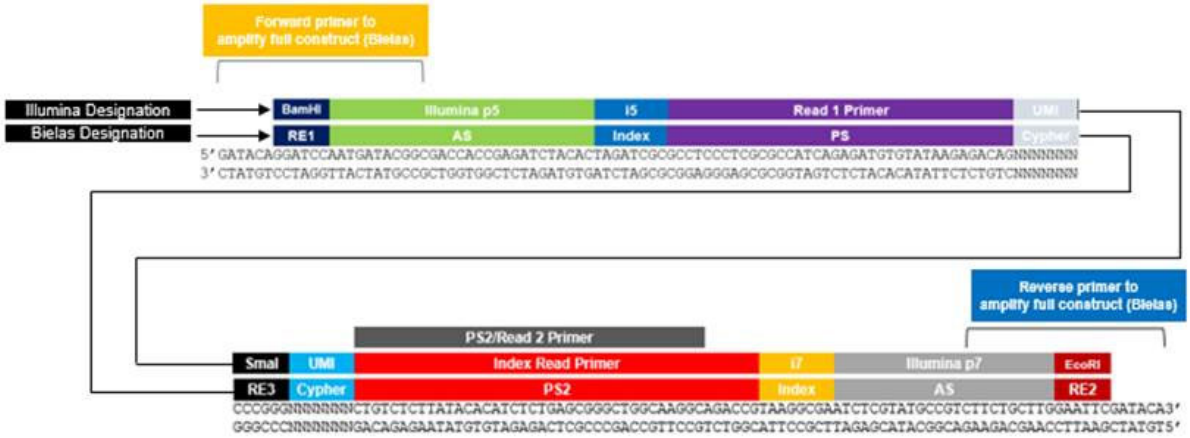


Fig. 2

607. Bielas explains that RE3 site is a restriction endonuclease recognition site that allowed Bielas to cut the vector with a restriction enzyme and insert the DNA fragments from the sample. EX1013 (Bielas) at 0023. This process places the identifier cite/Cypher at the junction between the adaptor and nucleic acid fragment as shown in the figure below.



608. In Example 3, Bielas provides the sequence of the adaptor. EX1013 (Bielas) at 0028–29. As shown in the annotated figure below, and as I further explained in Section VI.D.1.b of my declaration, the identifier site/Cypher is at the junction between the adaptor and nucleic acid fragment.



609. To the extent Bielas does not anticipate claim 10, Bielas renders it obvious based on Bielas's express teachings that the Cypher sequence is placed at the junction between the adaptor and nucleic acid fragment. EX1013 (Bielas) at 0028–29, 0041. Based on these teachings in Bielas, the POSA would have had a reasonable expectation of success that the methods in Bielas could be applied to the use of an adaptor with the identifier site at the junction between the adaptor and nucleic acid fragment.

4. Bielas Anticipates And/Or Renders Obvious The Challenged Claims Of The '241 Patent.

610. Bielas teaches every element of the method claimed in claims 1–16 of the '241 patent and thus anticipates the claims and/or renders the claims obvious. Specifically, Bielas teaches using a method using adaptors containing all of the features recited in the claims of the '241 patent, including an appended adaptor containing an identifier site (which Bielas refers to as a "Cypher"). EX1013 (Bielas) at 0041, 0038. Bielas further teaches using the Cypher sequence and sequence of the nucleic acid fragment to identify PCR copies of original fragments after an NGS read, which necessarily requires identifying a duplicate sequencing read. EX1013 (Bielas) at 0026–32.

a. Bielas Anticipates Claim 1.

611. For the reasons stated in Section VI.D.1 of my declaration and below

in this section, Bielas anticipates claim 1 of the ’241 patent.

1) “A method for detecting a duplicate sequence read from a population of sample sequencing reads”

612. The preamble to claim 1 of the ’241 patent, the preamble to claim 1 of the ’357 patent, and the preamble to claim 1 of the 399 patent are directed to the same method. The POSA would have understood that there was no substantive difference between the three preambles, even though the ’241 and ’357 preambles are plural (i.e., “detecting duplicate sequencing reads”) and the ’399 preamble is singular (i.e., “detecting a duplicate sequencing read”).

’241 Preamble	’357 Preamble	’399 Preamble
1. “A method for detecting duplicate sequencing reads” EX1005 at 25:33.	1. “A method for detecting duplicate sequencing reads” EX1004 at 25:36.	1. “A method for detecting a duplicate sequencing read from a population of sample sequencing reads” EX1003 at 25:37–38.

613. Because the ’241 claim 1 preamble is substantively the same as the ’357 and ’399 preambles that I discuss above, my explanation in Sections VI.D.2.a.1 and VI.D.3.a.1 applies equally here. Specifically, the POSA would have recognized the Bielas teaches a method of detecting and removing duplicate sequencing reads.

2) “obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor”

614. Claim 1 next recites “obtaining amplicons each comprising an

amplified fragment of a nucleic acid with an appended adaptor." EX1005 ('241 patent) at 25: 36–37. For the reasons I discuss below, Bielas discloses this limitation.

615. Bielas teaches appending adaptors with random tags to fragments prior to sequencing and then amplifying those labeled fragments, as I describe in Sections VI.D.1 and VI.D.2.a.2-3 of my declaration. Specifically, as I discuss in further detail in Section VI.D.1 above, both Example 1 and Example 3 ligate a nucleic acid fragment into an adaptor-containing vector, and then amplify the adaptor-linked fragments using PCR to obtain amplicons. EX1013 (Bielas) at 26–27; EX1013 (Bielas) at 28–32.

616. Ligation is a method of appending an adaptor. A POSA would not understand the word "appended" to place any limitation on the method that may be used to attach the adaptor to the fragment, and the '241 patent does not define the term.

617. Moreover, the '241 patent explains that "[i]n some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction." EX1005 ('241 patent) at 0016 at 9:34–37. Because this disclosure says it describes "some embodiments," the POSA would understand that it does not exclude other means of attaching the adaptor. But regardless, Bielas appended the adaptors via ligation, which is one of the two types of methods, in addition to priming reactions, of

"appending" specifically disclosed in the '241 patent. EX1005 ('241 patent) at 9:34–37.

3) "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment"

618. Claim 1 next recites "wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment." EX1005 ('241 patent) at 25:37–39. For the reasons I discuss above in connection with the '399 patent in Section VI.D.2.a.2.c and the '357 patent in Section VI.D.3.a.2, Bielas discloses this limitation.

619. As I discuss in further detail above in Section VI.D.1.a, in Example 1, Bielas ligated an adaptor with the structure shown in Bielas Figure 2. EX1013 (Bielas) at 0026-27, 0041. Figure 2 shows that the adaptors contained two Cyphers, each of which were a random nucleotide sequence unique to the amplified fragment. EX1013 (Bielas) at 0041.

620. As I discuss in further detail above in Section VI.D.1.b, in Example 3, Bielas provides the adaptor structure and sequence and specifies a 7-nucleotide Cypher on the 5' end of the nucleic acid fragment. EX1013 (Bielas) at 0028–29. As I describe above in Section VI.D.1.b of my declaration, Example 3 includes two Cyphers of "NNNNNNN," EX1013 (Bielas) at 0028–29, which, as I explained, the POSA would have understood to be the standard notation for random nucleotides.

Accordingly, the adaptor used in Example 3 contained a random nucleotide sequence unique to the amplified fragment.

4) "sequencing the amplicons to generate sequence reads that include identifier and target sequences; and"

621. Claim 1 next recites "sequencing the amplicons to generate sequence reads that include identifier and target sequences; and." EX1005 ('241 patent) at 25:40–41. Bielas discloses this limitation for the reasons I discuss above in Sections VI.D.1 and VI.D.2.a.4 of my declaration, as well as the reasons explained in this section.

622. This step in claim 1 of the '241 patent is substantively identical to the sequencing step in claim 1 of the '357 and '399 patents, as shown in the table below.

'241 Sequencing Step	'357 Sequencing Step	'399 Sequencing Step
"sequencing the amplicons to generate sequence reads that include identifier and target sequences" EX1005 at 25:40–41.	"sequencing the amplicons to produce sequence reads that include identifier and target sequences" EX1004 at 25:44–45.	"c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products" EX1003 at 25:44–51.

623. Despite the slight differences in the language, there is no substantive difference between the steps of the '241, '357, and '399 patents. The '241 patent uses the term "generate" and the '357 patent uses the term "produce." In this context, there is no substantive difference between the terms, and the POSA would have

understood "generate sequence reads" to mean the same as "produce sequence reads." Accordingly, this limitation is disclosed in Bielas for the reasons I discuss above in connection with Claim 1 of the '399 patent in Section VI.D.2.a.4 and Claim 1 of the '357 patent in Section VI.D.3.a.3.

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

624. Claim 1 next recites "identifying sequence reads with identical identifier and target sequences as duplicates." EX1005 ('241 patent) at 25:42–43.

625. This step in claim 1 of the '241 patent is the same as the fourth step in claim 1 of the '357 patent and substantively the same as the fourth step in claim 1 of the '399 patent, as shown in the table below.

'241 Identification Step	'357 Identification Step	'399 Identification Step
"identifying sequence reads with identical identifier and target sequences as duplicates" EX1004 at 25:46–47.	"identifying sequence reads with identical identifier and target sequences as duplicates" EX1004 at 25:46–47.	"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" EX1003 at 25:52–55.

626. In this context, the POSA would have understood that there was no substantive difference between the claim limitation in claim 1 of the '241 patent, claim 1 of the '357 patent, and claim 1 of the '399 patent. For the reasons I discuss

above with respect to claim 1 of the ’399 and claim 1 of the ’357 patents in Sections VI.D.2.a.5 and VI.D.3.a.4, respectively, the POSA also would have understood that Bielas teaches this limitation of the ’241 patent.

b. Claim 2: “The method of claim 1, further comprising removing a duplicate read from the sequence of reads.”

627. Claim 2 is anticipated by Bielas. Claim 2 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. Claim 2 of the ’241 patent recites the same limitations as claim 2 of the ’357 patent and claim 2 of the ’399 patent, as shown in the table below.

’241 Claim 2	’357 Claim 2	’399 Claim 2
2. “The method of claim 1, further comprising removing a duplicate read from the sequence reads.” EX1005 at 25:44–45	2. “The method of claim 1, further comprising removing a duplicate read from the sequence reads.” EX1004 at 25:48–49.	2. “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.” EX1003 at 25:61–64.

628. Because the limitation of claim 2 of the ’241 patent claim is substantively the same as the limitations of claim 2 of the ’399 patent and claim 2 of the ’357 patent, Bielas discloses this limitation for the reasons stated with respect to those claims in Sections VI.D.2.c and VI.D.3.c of my declaration.

c. Claim 3: "The method of claim 1, wherein the identifier site is sequenced with the target sequence."

629. Claim 3 is anticipated by Bielas for the reasons I describe above. Claim 3 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. Claim 3 of the '241 patent recites the same limitations as claim 3 of the '357 patent, as shown in the table below.

'241 Claim 3	'357 Claim 3
3. "The method of claim 1, wherein the identifier site is sequenced with the target sequence." EX1005 at 25: 46–48.	3. "The method of claim 1, wherein the identifier site is sequenced with the target sequence." EX1004 at 25:50–51.

630. Because Claim 3 of the '241 patent is substantively the same as claim 3 of the '357 patent, it is anticipated for the same reasons as Claim 3 of the '357 patent that I discuss above in Section VI.D.3.d.

d. Claim 4: "The method of claim 1, wherein the identifier site is sequenced separately from the target sequence."

631. Claim 4 is anticipated and/or obvious based on Bielas for the reasons I describe above. Claim 4 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a.

632. In addition to the limitations of claim 1, claim 4 of the '241 patent recites the same limitations as claim 4 of the '357 patent, as shown in the table below.

’241 Claim 4	’357 Claim 4
4. “The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.” EX1005 at 25: 48–49.	4. “The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.” EX1004 at 26:36–37.

633. Thus, Claim 4 of the ’241 patent is anticipated and/or obvious for the reasons I discuss above in connection with claim 4 of the ’357 patent in Section VI.D.3.e above.

e. Claim 5: “The method of claim 1, wherein the amplicons include amplified fragments of cDNA.”

634. Claim 5 is obvious based on Bielas. Claim 5 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. With respect to the additional limitation of Claim 5, both of Bielas’s examples that I discuss above in detail in Section VI.D.1 used genomic DNA. However, it would have been obvious to the POSA to modify Bielas’s methods for a library of cDNA. Bielas states multiple times that the methods and adaptors disclosed therein are suitable for use with cDNA specifically. EX1013 (Bielas) at 0007, 0014. Moreover, the POSA would have recognized at the priority date that there was no significant difference between the libraries of genomic DNA used in Bielas’s examples and a library of cDNA molecules, other than their respective origin. The POSA would have been motivated to use Bielas’s methods with cDNA because of the disclosed benefits of the methods, including permitting accurate molecular counting and

mitigating PCR and sequencing error. For these reasons, the POSA also would have had a reasonable expectation of success in practicing Bielas's methods with cDNA.

f. Claim 6: "The method of claim 1, wherein the amplicons include amplified fragments of genomic DNA."

635. Claim 6 is anticipated by Bielas. Claim 6 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. As I discuss in detail in Section VI.D.1, Example 1 and Example 3 in Bielas used genomic DNA as input material. EX1013 (Bielas) at 0026, 0030.

g. Claim 7: "The method of claim 1, wherein each adaptor includes a primer binding site."

636. Claim 7 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. Claim 7 is also anticipated because the additional limitation of a primer binding site in the adaptor is disclosed in Bielas. For the reasons I discuss above in Section VI.D.2.a.2, in connection with the primer binding sites recited in Claim 1 of the '399 patent, the adaptors used in both Example 1 and Example 3 had primer binding sites. EX1013 (Bielas) at 0026-31, 0041. As I discuss in further detail in Section VI.D.1, the adaptors were binding sites for multiple primers, including the primers Bielas used for PCR to copy the adaptor-ligated fragments out of the vectors, and the read primers and indexing primers in the Illumina devices used by Bielas. EX1013 (Bielas) at 0026-31, 0041.

h. Claim 8: "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site."

637. Claim 8 is anticipated by Bielas. Claim 8 depends from claim 7 and claim 7 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.g. Claim 8 of the '241 patent recites the same limitation as claim 7 of the '357 patent and claim 16 of the '399 patent, as shown in the table below.

'241 Claim 8	'357 Claim 7	'399 Claim 16
8. "The method of claim 7, wherein the primer binding site is a universal target sequencing primer binding site." EX1005 at 26:56–57.	7. "The method of claim 1, wherein the primer binding site is a universal target sequencing primer binding site." EX1004 at 42–43.	16. "The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site." EX 1003 at 65–67.

638. Thus, Claim 8 of the '241 patent is anticipated for the reasons I discuss above in connection with claim 16 of the '399 patent in Section VI.D.2.n, and claim 7 of the '357 patent in Section VI.D.3.g above.

i. Claim 9: "The method of claim 1, wherein each unique identifier is at least 6 nucleotides in length."

639. Claim 9 is anticipated by Bielas. Claim 9 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. As I discuss in Section VI.D.1, in both Example 1 and Example 3, Bielas disclosed adaptors with Cyphers. In Example 1, as I discuss above in Section VI.D.2.c in

connection with claim 1 of the ’399 patent, Bielas disclosed that the Cypher could be 6, 7, or 8 nucleotides in length. Moreover, Bielas disclosed ranges of Cypher lengths, including 7 to 9 nucleotides that overlap with the range of at least 6 recited in claim 9 of the ’241 patent.

640. In Example 3, as I discuss above in Section VI.D.2.c in connection with claim 1 of the ’399 patent, Bielas disclosed an adaptor with two Cyphers of 7 nucleotides. EX1013 (Bielas) at 0028-29.

j. Claim 10: “The method of claim 1, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction.”

641. Claim 10 is anticipated by Bielas. Claim 10 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. For the reasons I explain above in Section VI.D.4.a.2, Bielas teaches appending adaptors to fragments using a ligation reaction. EX1013 (Bielas) at 0026-27, 0028-31.

k. Claim 11: “The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.”

642. Claim 11 is anticipated by and/or obvious over Bielas. Claim 11 depends from claim 1 and Bielas anticipates Claim 1 for the reasons I describe above in Section VI.D.4.a. In addition, Claim 11 of the ’241 patent recites the same limitations as claim 10 of the ’357 patent, as shown in the table below.

’241 Claim 11	’357 Claim 10
11. “The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.” EX1005 at 26:35–37.	10. “The method of claim 1, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.” EX1004 at 26:48–50.

643. For the reasons I discuss in connection with claim 10 of the ’357 patent in Section VI.D.3.i, Bielas anticipates and/or renders obvious claim 11 of the ’241 patent. EX1013 (Bielas) at 0028-29, 0041.

l. Claim 12: “The method of claim 1, further comprising determining the genomic start position for each sequencing read.”

644. Bielas anticipates and/or renders obvious claim 12. Claim 12 depends from claim 1 and Bielas anticipates the limitations of claim 1 for the reasons I explain in Section VI.D.4.a. With respect to the additional limitations of claim 12, Bielas expressly teaches determining the start position for each read.

645. Specifically, Bielas explains in Example 1 that “unique cypher tags are used to computationally deconvolute the sequencing data and map all sequencing reads to single molecules” and that “[b]ase calling and sequence alignment will be performed.” EX1013 (Bielas) 0026–27. Sequence alignment expressly refers to mapping the observed fragment reads against a known reference sequence (in Example 1, the human genome) and thereby determining the genomic start position for each read.

646. Similarly, in Example 3, Bielas sequenced fragments of the TP53 gene.

EX1013 (Bielas) at 0029. As part of the data analysis of the fragments, Bielas explains that "[s]equences were then compared to wild-type TP53 sequence." EX1013 (Bielas) at 0006. To perform this analysis, Bielas necessarily aligned the observed fragments to the known TP53 gene sequence and thereby determined the genomic start position.

647. To the extent Bielas does not anticipate this claim, claim 12 would have been obvious to the POSA. The POSA would have recognized that, as taught in Example 1 of Bielas, EX1013 (Bielas) at 0027, aligning sequence reads to a reference genome was a standard part of data analysis at the priority date. This is particular true in the context of Bielas's emphasis on "true mutations," which are mutations from the germline reference genome, EX1013 (Bielas) at 0006, as opposed to artifact mutations created by PCR or sequencing error. For these reasons, the POSA would have had a reasonable expectation of success in practicing the method of claim 12 based on the teachings of Bielas.

m. Claim 13: "The method of claim 1, wherein each adaptor further comprises an indexing site."

648. Bielas anticipates claim 13. Claim 13 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. For the reasons I discuss above in connection with claim 1 of the '399 patent in Section VI.D.2.a.2.b, the adaptors used in both Example 1 and Example 3 included indexing

sites/sample indexes. EX1013 (Bielas) at 0027, 0028-29, 0041.

n. Claim 14: "The method of claim 1, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length."

649. Claim 14 is anticipated by Bielas. Claim 14 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. Claim 14 recites an identifier site of 2-10 nucleotides. For the reasons I discuss above in Section VI.D.2.a.2.c in connection with the range of 1-8 nucleotides in claim 1 of the '399 patent, Example 1 discloses this limitation by disclosing Cyphers of 6, 7, and 8 nucleotides. EX1013 (Bielas) at 0015. Likewise, Example 3 in Bielas discloses this limitation in its use of a Cypher of 7 nucleotides. EX1013 (Bielas) at 0028–29.

o. Claim 15: "The method of claim 1, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons."

650. Claim 15 is anticipated by and/or obvious based on Bielas. Claim 15 depends from claim 1 and claim 1 is anticipated by Bielas for the reasons I discuss above in Section VI.D.4.a. As I explain further above in section VI.D.1 of my declaration, Bielas disclosed each of these steps as part of Example 1. EX1013 (Bielas) at 0026–27, 0028–32 In Example 1, Bielas obtained samples of genomic

DNA, fragmented the samples, ligated them to adaptor-containing vectors, and then amplified the adaptor-ligated fragments. EX1013 (Bielas) at 0026-27.

651. To the extent that Bielas does not anticipate these limitations, it renders them obvious as part of the methods described in Bielas. Bielas expressly describes that it can be used with "double-stranded nucleic acid molecules that may be fragments or shorter molecules generated from longer nucleic acid molecules" and that "[t]arget nucleic acid fragments from longer molecules may be generated using a variety of techniques known in the art, such as mechanical shearing or specific cleavage with restriction endonucleases." EX1013 (Bielas) at 0010–11. As I discuss above in Section VI.D.1, Bielas taught the other limitations of claim 15, including appending adaptors and an amplification reaction as part of both Examples 1 and 3. For these reasons, the POSA also would have had a reasonable expectation of success in practicing the method of claim 15 based on Bielas.

p. Claim 16: "The method of claim 1, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index."

652. Claim 16 would have been anticipated by and/or obvious based on Bielas. Claim 16 depends from claim 1 and claim 1 is anticipated by Bielas for the

reasons I discuss above in Section VI.D.4.a. Claim 16 recites a routine implementation for sample multiplexing. Bielas anticipates this limitation in explaining that the "Index" is a sample index used for sample pooling "if multiplex sequencing is desired." EX1013 (Bielas) at 0022, 0027. As I describe above in Section VI.D.1. of my declaration, Bielas discloses "a unique index sequence (comprising a length ranging from about 4 nucleotides to about 25 nucleotides) specific for a particular sample is included with each dual cypher vector library. For example, if ten different samples are being pooled in preparation for multiplex sequencing, then ten different index sequences will be used such that ten dual cypher vector libraries are used in which each library has a single, unique index sequence identifier (but each library has a plurality of random cyphers)." EX1013 (Bielas) at 0015–16. The POSA would understand this teaching regarding sample indexes was consistent with the well-known use of sample indexes in NGS. Thus, the POSA would understand Bielas to disclose methods of pooling at least two samples (or more) wherein each sample had a unique sample index, and then separating the resulting reads based on those sample indexes.

653. To the extent Bielas does not anticipate this claim, at a minimum, this claim would have been obvious to the POSA based on Bielas. Bielas expressly discloses sample multiplexing. EX1013 (Bielas) at 0015-16, 0022, 0027. And the pooling of samples for multiplexing using sample indexes was routine at the priority

date as I discuss above in Section IV.K.2.a. The POSA would have recognized that the simplest implementation of this concept was to pool two samples wherein each has a different sample index. Moreover, separation of the samples was a standard part of Illumina workflows at the priority date. In view of Bielas's express disclosure of sample multiplexing, the POSA would have had a reasonable expectation of success in practicing the method of claim 16.

VII. PRIOR ART CONSIDERED DURING PROSECUTION

654. I understand from counsel that the Patent Trial and Appeals Board (the "Board") in determining whether to institute IDT's petitions for IPR will consider, among other things, whether prior art and arguments presented to the Examiner during prosecution were the same or substantially similar to the prior art and arguments raised in IDT's petitions. I further understand that the Board may also consider whether the Examiner clearly erred in allowing the claims.

655. I have reviewed the file histories of the '399, '357, and '241 patents. None of the prior art I rely on in this declaration to form the bases of my opinions—namely, EX1011 (Iafrate), EX1012 (Kivioja), and EX1013 (Bielas)—were cited by the Examiner. I have reviewed the art relied on by the Examiner in rejecting the claims of the patents during prosecution including the McCloskey publication (EX1022) and the Samuels patent application (EX1033), as well as the Examiner's and Tecan/NuGEN's statements about those references. In my opinion, neither

Iafrate, Kivioja, nor Bielas are substantially similar to McCloskey or Samuels, which teach materially different methods. As relevant in particular to the issued claims of the '399, '357, and '241 patents, for example, as I explained above in Sections VI.C.4, VI.C.6.a, VI.C.7.a, VI.D.1, VI.D.2.a, and VI.D.3.a, Kivioja and Bielas both taught using the molecular tag sequence and the fragment sequence together to identify duplicate sequencing reads.

656. To the extent the Board concludes that McCloskey and Samuels are substantially similar to Iafrate, Kivioja, and/or Bielas, in my opinion, the Examiner necessarily erred because, as explained in my opinions above, Iafrate, Kivioja, and Bielas anticipate and/or render the claims obvious.

VIII. CONCLUSION

657. For the reasons discussed above, it is my opinion that the Challenged Claims are invalid.

I declare under penalty of perjury that the foregoing is true and correct. I also hereby declare that all statements made herein on my knowledge are true and that all statements made on information and belief are believed to be true, and further, that those statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: September 30, 2024



Peter A. Sims, Ph.D.

EXHIBIT A

Peter A. Sims, Ph.D.

Columbia University Irving Medical Center
3960 Broadway
Lasker Biomedical Research Building, Room 203AC
New York, NY 10032
(212) 305-2687
pas2182@cumc.columbia.edu

Professional Appointments

Member, Planning Committee, Chan Zuckerberg Biohub New York, 2023-present

Co-Leader, Precision Oncology and Systems Biology Program, Herbert Irving Comprehensive Cancer Center, Vagelos College of Physicians & Surgeons, Columbia University, 2023-present

Scientific Director, Sulzberger Columbia Genome Center, Vagelos College of Physicians & Surgeons, Columbia University, 2022-present

Associate Professor (with tenure), Department of Systems Biology and Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians & Surgeons, Columbia University, 2020-present

Faculty Director, Columbia Single Cell Analysis Core, Department of Systems Biology, Vagelos College of Physicians & Surgeons, Columbia University, 2017-present

Director, Systems Biology Graduate Studies, Department of Systems Biology, Vagelos College of Physicians & Surgeons, Columbia University, 2016-2022

Assistant Professor, Department of Systems Biology and Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians & Surgeons, Columbia University, 2012-2020

Education and Training

Postdoctoral Fellow, Harvard University 2009-2011
Cambridge, MA

Ph.D. Chemistry, Harvard University, 2009
Cambridge, MA

A.M. Chemistry, Harvard University, 2006
Cambridge, MA

B.S. Chemistry (*summa cum laude*), University of California San Diego, 2004
La Jolla, CA

Awards and Honors

Research Initiatives for Science and Engineering (RISE) Award, Columbia University, 2021
Emerging Leader Award, The Mark Foundation for Cancer Research, 2018
Research Initiatives for Science and Engineering (RISE) Award, Columbia University, 2014
Cold Spring Harbor Asia Fellowship, 2011
Biophysical Society Student Travel Award, 2008
American Chemical Society Contributed Talk Award, 2007
Delegate to the Annual Meeting of Nobel Laureates in Lindau, U.S. Dept. of Energy, 2006
Certificate of Distinction in Teaching, Harvard University, 2005, 2006
National Science Foundation Graduate Research Fellowship, 2004-2007
Joseph E. Mayer Award for Undergraduate Research, UCSD, 2004
Teaching Assistant Excellence Award, UCSD, 2004
Phi Beta Kappa, UCSD, 2003
Barry M. Goldwater Scholarship, 2003-2004

Funding

NIH/NIBIB K01 Career Development Award, PI, 2012-2017
NIH Common Fund R21, Single Cell Analysis Program, PI, 2012-2014
Simons Foundation, SFARI Explorer Award, PI, 2012-2013
Irving Institute for Clinical and Translational Research, CaMPR-BASIC Award, Co-PI, 2013-2014
Brain Tumor Ecology Collaborative Subaward, James S. McDonnell Foundation, Co-I, 2013-2015
NHI/NCI U54 Physical Sciences Oncology Center, Co-I, 2015-2020
NIH/NCI R33 Innovative Molecular Analysis Technologies Award, PI, 2016-2020
NIH/NCI U54 Center for Cancer Systems Biology, Core Director, 2016-2021
NIH/NIAID U19 Human Immunology Project Consortium, Core Director, 2016-2021
NIH/NINDS R01 Research Project Grant, Co-PI, 2017-2022
Chan Zuckerberg Initiative Human Cell Atlas Pilot Grant, PI, 2017-2018
NIH/NIDDK P20 Developmental Center Grant, Co-PI, 2018-2020
NIH/NIAID P01 Program Project Grant, Core Director, 2018-2023
NIH/NHLBI U01 Immunobiology and alveolar physiology of the aging lung, Co-I, 2018-2023
Mark Foundation Emerging Leader Award, PI, 2018-2022
Chan Zuckerberg Initiative Seed Networks for the Human Cell Atlas, Co-PI, 2019-2023
NIH/NCI R01 Cancer patient on a chip, Co-I, 2020-2025
Aging Biology Foundation, Single-cell analysis of aging in lymphoid compartments, PI, 2021-2024
NIDDK P30 Digestive and Liver Disease Research Center, Core Director, 2022-2027
NIH/NIAID U19 Human Immunology Project Consortium, Co-Program Director, 2022-2027
DARPA Assessing Immune Memory Program, Co-PI, 2022-2026
NIH/NINDS R01 Research Project Grant, Co-PI, 2023-2028
NIH/NCI U54 Center for Cancer Systems Biology (Columbia), Core Director, 2023-2028
NIH/NCI U54 Center for Cancer Systems Biology (Mayo Clinic), Co-I, 2023-2028
Chan Zuckerberg Biohub Investigator Award, PI, 2024-2029

Publications

* equal contribution; † co-corresponding author

1. Li, A.L., Sugiura, K., Nishiwaki, N., Suzuki, K., Sadeghian, D., Zhao, J., Maitra, A., Falvo, D., Chandwani, R., Pitarresi, J.R.†, **Sims, P.A.**†, Rustgi, A.K.† Mapping cellular plasticity through the lens of FRA1 during $Kras^{G12D}$ mediated pancreatic acinar to ductal metaplasia, *Developmental Cell*, in press (2024).

2. Banu, M.*, Dovas, A.*, Argenziano, M.G.*, Zhao, W.*, Sperring, C.P., Grajal, H.C., Liu, Z., Higgins, D.M.O., Amini, M., Pereira, B., Ye, L.F., Mahajan, A., Humala, N., Furnari, J.L., Upadhyayula, P.S., Zandkarimi, F., Nguyen, T.T.T., Teasley, D., Wu, P.B., Li, H., Karan, C., Dowdy, T., Razavilar, A., Siegelin, M.D., Kitajewski, J., Larion, M., Bruce, J.N, Stockwell, B.R., **Sims, P.A.**†, Canoll, P.† A cell state metabolic vulnerability to GPX4-dependent ferroptosis in glioblastoma, *EMBO Journal*, in press (2024).

3. Capdevila, C., Miller, J., Cheng, L., Kornberg, A., George, J.J., Lee, H., Botella, T., Moon, C.S., Murray, J.W., Lam, S., Calderon, R.I., Malagola, E., Whelen, G., Lin, C.-S., Han, A., Wang, T.C., **Sims, P.A.**, Yan, K.S. Time-resolved fate mapping identifies the intestinal upper crypt zone as an origin of Lgr5+ crypt base columnar cells. *Cell*, 187, 3039 (2024).

4. Gray, J.I., Caron, D.P., Wells, S.B., Guyer, R., Szabo, P., Rainbow, D., Ergen, C., Rybkina, K., Bradley, M.C., Matsumoto, R., Pethe, K., Kubota, M., Teichmann, S., Jones, J., Yosef, N., Atkinson, M., Brusko, M., Brusko, T.M., Connors, T.J., **Sims, P.A.**, Farber, D.L. Human γδ T cells in diverse tissues exhibit site-specific maturation dynamics across the lifespan. *Science Immunology*, 9, eadn3954 (2024).

5. Tuddenham, J.F., Taga, M., Haage, V., Roostaei, T., White, C., Lee, A., Fujita, M., Hairallah, A.K., Green, G., Hyman. B., Frosch, M., Hopp, S., Beach, T.G., Corboy, J., Habib, N., Klein, H.-U., Soni, R.K., Teich, A.F., Hickman, R.A., Alcalay R.N., Shneider, N., Schneider, J., **Sims, P.A.**, Bennett, D.A., Olah, M., Menon, V., De Jager, P.L. A cross-disease human microglial framework identifies disease-enriched subsets and tool compounds for microglial polarization. *Nature Neuroscience*, in press (2024).

6. Cheng, Y.L., Banu, M.A., Zhao, W., Rosenfeld, S.S., Canoll, P., **Sims, P.A.** Multiplexed single-cell lineage tracing of mitotic kinesin inhibitor resistance in glioblastoma. *Cell Reports*, 43, 114139 (2024).

7. Mi, X., Bekerman, W., Rustgi, A.K., **Sims, P.A.**, Canoll, P.D., Hu, J. RZiMM-scRNA: A regularized zero-inflated mixture model framework for single-cell RNA-seq data. *Annals of Applied Statistics*, 18, 1 (2024).

8. Liu, Z., Mela, A., Argenziano, M., Banu, M., Furnani, J., Kotidis, C., Sperring, C., Humala, N., Mahajan, A., Bruce, J.N., Canoll, P., **Sims, P.A.** Single-Cell Analysis of 5-ALA Intraoperative Labeling Specificity for Glioblastoma. *Journal of Neurosurgery*, 140, 968 (2023).

9. Matsumoto, R.*, Gray, J.*, Rybkina, K., Oppenheimer, H., Levy, L., Friedman, L.M., Khamaisi, M., Meng, W., Rosenfeld, A.M., Guyer, R.S., Bradley, M.C., Chen, D., Atkinson, M.A., Brusko, T.M., Brusko, M., Connors, T.J., Luning Prak, E.T., Hershberg, U., **Sims, P.A.**, Hertz, T., Farber, D.L. Induction of bronchus-associated lymphoid tissue is an early life adaptation for promoting human B cell immunity. *Nature Immunology*, 24, 1370 (2023).

10. Connors, T.J., Matsumoto, R., Verma, S., Szabo, P.A., Guyer, R., Gray, J., Wang, Z., Thapa, P., Dogra, P., Poon, M.M.L., Rybkina, K., Bradley, M.C., Idzikowski, E., McNichols, J., Kubota, M., Pethe K., Shen, Y., Atkinson, M.A., Brusko, M., Brusko, T.M., Yates, A.J., **Sims, P.A.**, Farber, D.L. Site-specific development and progressive maturation of human tissue-resident memory T cells over infancy and childhood. *Immunity*, 56, 1894 (2023).

11. Sugiura, K., Masuike, Y., Suzuki, K., Shin, A.E., Sakai, N., Matsubara, H., Ohtsuka, M., **Sims, P.A.**, Lengner, C.J., Rustgi, A.K. LIN28B promotes cell invasion and colorectal cancer metastasis via CLDN1 and NOTCH3. *JCI Insight*, e167310, (2023).

12. Al-Dalahmah, O.†, Argenziano, M., Kannan, A., Mahajan, A., Furnari, J., Paryani, F., Boyett, D., Save, A., Humala, N., Khan, F., Li, J., Lu, H., Sun, Y., Tuddenham, J.F., Goldberg, A.R., Dovas, A., Banu, M.A., Sudhakar, T., Bush, E., Lassman, A.B., McKhann, G.M., Gill, B.J.A., Youngerman, B., Sisti, M.B., Bruce, J.N., **Sims, P.A.**, Menon, V.†, Canoll, P.† Re-convolving the compositional landscape of primary and recurrent glioblastoma reveals prognostic and targetable tissue states. *Nature Communications*, 14, 2586 (2023).

13. Hobson, B.D., Stanley, A.T., De Los Santos, M.B., Culbertson, B., Mosharov, E.V, **Sims, P.A.**†, Sulzer, D.† Conserved and cell type-specific transcriptional responses to IFN-g in the ventral midbrain. *Brain, Behavior, and Immunity*, 111, 277 (2023).

14. Mundi, PS, Dela Cruz, FS, Grunn, A, Diolaiti, D, Mauguen, A, Rainey, AR, Guillan, K, Siddiquee, A, You, D, Realubit, R, Karan, C, Ortiz, MV, Douglass, EF, Accordino, M, Mistretta, S, Brogan, F, Bruce, JN, Caescu, CI, Carvajal, RD, Crew, KD, Decastro, G, Heaney, M, Henick, BS, Hershman, DL, Hou, JY, Iwamoto, FM, Jurcic, JG, Kiran, RP, Kluger, MD, Kreisl, T, Lamanna, N, Lassman, AB, Lim, EA, Manji, GA, McKhann, GM, McKiernan, JM, Neugut, AI, Olive, KP, Rosenblat, T, Schwartz, GK, Shu, CA, Sisti, MB, Tergas, A, Vattakalam, RM, Welch, M, Wenske, S, Wright, JD, Canoll, P, Hibshoosh, H, Kalinsky, K, Aburi, M, **Sims, PA**, Alvarez, MJ, Kung, AL, Califano, A. A transcriptome-based precision oncology platform for patient-therapy alignment in a diverse set of treatment resistant malignancies. *Cancer Discovery*, CD-22-1020 (2023).

15. Poon, M.M.L.*, Caron, D.P.*, Wang, Z., Wells, S.B., Chen, D., Meng, W., Szabo, P.A., Lam, N., Kubota, M., Matsumoto, R., Rahman, A., Prak, E.T.L., Shen, Y., **Sims. P.A.**, Farber, D.L. Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nature Immunology*, 24, 309 (2023).

16. Spinazzi, EF, Argenziano, MG, Upadhyayula, PS, Banu, MA, Neira, JA, Higgins, DMO, Wu, PB, Pereira, B, Mahajan, A, Humala, N, Al-Dalahmah, O, Zhao, W, Save, AV, Gill, BJA, Boyett, DM, Marie, T, Furnari, JL, Sudhakar, TD, Stopka, SA, Regan, MS, Catania, V, Good, L, Zacharoulis, S, Behl, M, Petridis, P, Jambawalikar, S, Mintz, A, Lignelli, A, Agar, NYR, **Sims**, **PA**, Welch, MR, Lassman, AB, Iwamoto, FM, D'Amico, RS, Grinband, J, Canoll, P, Bruce, JN. Chronic convection-enhanced delivery of topotecan for patients with recurrent glioblastoma: a first-in-patient, single-centre, single-arm, phase 1b trial. *Lancet Oncology*, 23, 1409 (2022).

17. Kenchappa, R.S., Dovas, A., Argenziano, M.G., Meyer, C.T., Stopfer, L.E., Banu, M.A., Pereira, B., Griffith, J., Mohammad, A., Talele, S., Haddock, A., Zarco, N., Elmquist, W., White, F., Quaranta, V., **Sims, P.**, Canoll, P., Rosenfeld, S.S. Activation of STAT3 through combined SRC and EGFR signaling drives resistance to a mitotic kinesin inhibitor in glioblastoma. *Cell Reports*, 39, 110991 (2022).

18. Dominguez Conde, C.*, Xu, C.*, Jarvis, L.*, Rainbow, D.*, Wells, S.B.*, Gomes, T., Howlett, S., Suchanek, O., Polanski, K., King, H., Mamanova, L., Huang, N., Szabo, P., Richardson, L., Bolt, L., Fasouli, E.S., Mahbubani, K., Prete, M., Tuck, E., Richoz, N., Tuong, Z.K., Campos, L., Mousa, H., Needham, E., Pritchard, S., Li, T., Elmentaite, R., Park, J.E., Rahmani, E., Chen, D., Menon, D., Bayraktar, O., James, L., Meyer, K., Yosef, N., Clatworthy, M., **Sims, P.A.**, Farber, D., Saeb-Parsy, K., Jones, J., Teichmann, S. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376, eabl5197 (2022).

19. Chait, M., Yilmaz, M.M., Shakil, S., Ku, A.W., Dogra, P., Connors, T.J., Szabo, P.A., Gray, J.I., Wells, S.B., Kubota, M., Matsumoto, R., Poon, M.M.I., Snyder, M.E., Baldwin, M.R., **Sims, P.A.**, Saqi, A., Farber, D.L., Weisberg, S.P. Immune and epithelial determinants of age-related risk and alveolar injury in fatal COVID-19. *JCI Insight*, e157608 (2022).

20. Metz, J.B.*, Hornstein, N.J.*, Das Sharma, S., Worley, J., Gonzalez, C., **Sims, P.A.** High-throughput translational profiling with riboPLATE-seq. *Scientific Reports*, 12, 5718 (2022).

21. Snyder, M.E., Moghbeli, K., Bondonese, A., Craig, A., Popescu, I., Fan, L., Tabib, T., Lafyatis, R., Chen, K., Bittar, H.E.T., Lendermon, E., Pilewski, J., Johnson, B., Kilaru, S., Zhang, Y., Sanchez, P.G., Alder, J.K., **Sims, P.A.**, McDyer, J.F. Modulation of tissue resident memory T cells by glucocorticoids after acute cellular rejection in lung transplantation. *Journal of Experimental Medicine*, 219, e20212059 (2022).

22. Masuda, K., Kornberg, A., Miller, J., Lin, S., Suek, N., Botella, T., Secener, K.A., Bacarella, A.M., Cheng, L., Ingham, M., Rosario, V., Al-Masrou, A.M., Lee-Kong, S.A., Kiran, R.P., Stoeckius, M., Smibert, P., Del Portillo, A., Oberstein, P.E., **Sims, P.A.**, Yan, K.S., Han, A., Multiplexed single-cell analysis reveals prognostic and non-prognostic T cell types in human colorectal cancer. *JCI Insight,* 7, e154646 (2022).

23. Hobson, B.D., Choi, S.J., Mosharov, E.V., Soni, R.K., Sulzer, D.†, **Sims, P.A.**† Subcellular proteomics of dopamine neurons in the mouse brain. *eLife*, 11, e70921 (2022).

24. Agrafioti, P., Morin-Baxter, J., Tanagala, K.K.K., Dubey, S., **Sims, P.**, Lalla, E., Momen-Heravi, F. Decoding the role of macrophages in periodontitis and type 2 diabetes using single-cell RNA-sequencing. *FASEB Journal*, 36, e22136 (2022).

25. Hobson, B.D., Kong, L., Angelo, M.F., Lieberman, O.J., Mosharov, E.V., Herzog, E.†, Sulzer, D.†, **Sims, P.A.**† Subcellular and regional localization of mRNA translation in midbrain dopamine neurons. *Cell Reports*, 38, 110208 (2022).

26. Bayin, N.S., Mizrak, D., Stephen, D.N., Lao, Z., **Sims, P.A.**, Joyner, A.L. Injury-induced ASCL1 expression orchestrates a transitory cell state required for repair of the neonatal cerebellum. *Science Advances*, 7, 10.1126/sciadv.abj598 (2021).

27. Poon, M.M.L, Byington, E., Meng, W., Kubota, M., Matsumoto, R., Grifoni, A., Weiskopf, D., Dogra, P., Lam, N., Szabo, P.A., Ural, B.B., Wells, S.B., Rosenfeld, A.M., Brusko, M.A., Brusko, T.M., Connors, T.J., Sette, A., **Sims, P.A.**, Prak, E.T.L, Shen, Y., Farber, D.L. Heterogeneity of human anti-viral immunity shaped by virus, tissue, age, and sex. *Cell Reports*, 37, 110071 (2021).

28. Arrieta, V.A., Chen, A.X., Kane, J.R., Kang, S.J., Kassab, C., Dmello, C., Zhao, J., Burdett, K.B., Upadhyayula, P.S., Lee-Chang, C., Shilati, J., Jaishankar, D., Chen, L., Gould, A., Zhang, D., Yuan, J., Zhao, W., Ling, X., Burks, J.K., Laffleur, B., Amidei, C., Bruce, J.N., Lukas, R.V, Yamaguchi, J.T., Cieremans, D., Rothschild, G., Basu, U., McCord, M., Brat, D.J., Zhang, H., Cooper, L.A.D, Zhang, B., **Sims, P.**, Cloughesy, T.F., Prins, R., Canoll, P., Stupp, R., Heimberger, A.B, Horbinski, C., Iwamoto, F.M. †, Rabadan, R. †, Sonabend, A.M. † ERK1/2 phosphorylation predicts survival following anti-PD-1 immunotherapy in recurrent glioblastoma. *Nature Cancer*, 2, 1372, (2021).

29. Kenchappa, R.S., Liu, Y., Argenziano, M.G., Banu, M.A., Mladek, A.C., West, R., Luu, A., Quinones-Hinojosa, A., Hambardzumyan, D., Justilien, V., Leitges, M., Sarkaria, J.N., **Sims, P.A.**, Canoll, P., Murray, N.R.†, Fields, A.P.†, Rosenfeld, S.S.† Protein kinase C_I and SRC signaling define reciprocally related subgroups of glioblastoma with distinct therapeutic vulnerabilities, *Cell Reports,* 37, 110054 (2021).

30. Capdevila, C., Trifas, M., Miller, J., Anderson, T., **Sims, P.A.**, Yan, K.S. Cellular origins and lineage relationships of the intestinal epithelium. *American Journal of Physiology: Gastrointestinal and Liver Physiology*, 321, G413 (2021).

31. Chen, A.X., Gartrell, R.D., Zhao, J., Upadhyayula, P.S., Zhao, W., Yuan, J., Minns, H.E., Dovas, A., Bruce, J.N., Lasorella, A., Iavarone, A., Canoll, P., **Sims, P.A.**, Rabadan, R. Single-cell characterization of macrophages in glioblastoma reveals MARCO as a mesenchymal pro-tumor marker. *Genome Medicine*, 13, 88 (2021).

32. Zhao, W.*, Dovas, A.*, Spinazzi, E.F.*, Levitin, H.M., Banu, M.A., Upadhyayula, P., Sudhakar, T., Marie, T., Otten, M.L., Sisti, M., Bruce, J.N., Canoll, P.†, **Sims, P.A.** † Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. *Genome Medicine*, 13, 82 (2021).

33. Affo, S., Nair, A., Brundu, F., Ravichandra, A., Bhattacharjee, S., Matsuda, M., Chin, L., Filliol, A., Wen, W., Song, X., Decker, A., Worley, J., Caviglia, J.M., Yu, L., Yin, D., Saito, Y., Savage, T., Wells, R.G., Mack, M., Zender, L., Arpaia, N., Remotti, H.E., Rabadan, R., **Sims, P.**, Leblond, A.-L., Weber, A., Riener, M.-O., Stockwell, B.R., Glaublomme, J., Llovet, J.M., Michalopoulos, G.K., Seki, E., Sia, D., Chen, X., Califano, A., Schwabe, R.F. Promotion of cholangiocarcinoma growth by diverse cancer-associated fibroblast subpopulations, *Cancer Cell*, 39, 866 (2021).

34. Bhattacharjee, S., Hamberger, F., Ravichandra, A., Miller, M., Nair, A., Affo, S., Filliol, A., Chin, L., Savage, T.M., Yin, D., Wirsik, N.M., Mehal, A., Arpaia, N., Seki, E., Mack, M., Zhu, D., **Sims, P.A.**, Stanger, B.Z., Olive, K.P., Schmidt, T., Wells, R.G., Mederacke, I., Schwabe, R.F. Tumor restriction by type I collagen opposes tumor-promoting effects of cancer-associated fibroblasts. *Journal of Clinical Investigation*, 146987 (2021).

35. Yun, T.J., Igarashi, S., Zhao, H., Perez, O.A., Pereira, M.R., Zorn, E., Shen, Y., Goodrum, F., Rahman, A., **Sims, P.A.**, Farber, D.L., Reizis, B. Human plasmacytoid dendritic cells mount a distinct antiviral response to virus-infected cells. *Science Immunology*, 6, eabc7302 (2021).

36. Szabo, P.A.*, Dogra, P.*, Gray, J.I.*, Wells, S.B.*, Connors, T.J., Weisberg, S.P., Krupska, I., Matsumoto, R., Poon, M.M.L., Idzikowski, E., Morris, S., Pasin, C., Yates, A.J., Ku, A., Chait, M., Davis-Porada, J., Guo, X.V., Zhou, J., Steinle, M., Mackay, S., Saqi, A., Baldwin, M., **Sims, P.A.**, Farber, D.L. Longitudinal

profiling of respiratory and systemic immune responses reveals myeloid cell-driven lung inflammation in severe COVID-19. *Immunity*, 54, 1 (2021).

37. Damaghi, M., West, J., Robertson-Tessi, M., Xu, L., Ferrall-Fairbanks, M.C., Stewart, P.A., Persi, E.A., Fridley, L.F., Altrock, P.M., Gatenby, R., **Sims, P.A.**, Anderson, A.R., Gillies, R.J. The harsh microenvironment in early breast cancer selects for a Warburg phenotype. *Proc. Natl. Acad. Sci USA*, 118, e2011342118 (2021).

38. Garofano, L., Migliozzi, S., Oh, Y.T., D'Angelo, F., Najac, R.D., Ko, A., Frangaj, B., Caruso, F.P., Yu, K., Yuan, J., Zhao, W., Di Stefano, A.L., Bielle, F., Jiang, T., **Sims, P.A.**, Suva, M.L., Tang, F., Su, X., Ceccarelli, M., Sanson, M., Lasorella, A, Iavarone, A. Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nature Cancer*, 2, 141 (2021).

39. Liu, Z.*, Yuan, J.*, Lasorella, A., Iavarone, A., Bruce, J.N., Canoll, P., **Sims, P.A.** Integrating single-cell RNA-seq and imaging with SCOPE-seq2. *Scientific Reports*, 19482 (2020).

40. Arriaga, J.M., Panja, S., Alshalalfa, M., Zhao, J., Zou, M., Giacobbe, A., Madubata, C.J., Kim, J.Y., Rodriguez, A., Coleman, I., Virk, R.K., Hibshoosh, H., Ertunc, O., Ozbek, B., Fountain, J., Karnes, R.J., Luo, J., Antonarakis, E.S., Nelson, P.S., Feng, F.Y., Rubin, M.A., De Marzo, A.M, Rabadan, R., **Sims, P.A.**, Mitrofanova, A. †, Abate-Shen, C. † A MYC and RAS co-activation signature in localized prostate cancer drives bone metastasis and castration resistance. *Nature Cancer*, 1, 1082 (2020).

41. Caruso, F., Garofano, L., D'Angelo, F., Yu, K., Tang, F., Yuan, J., Zhang, J., Cerulo, L., Pagnotta, S.M., Bedognetti, D., **Sims, P.A.**, Suva, M., Su, X.D., Lasorella, A., Iavarone, A., Ceccarelli, M. A map of tumor-host interactions in glioma at single-cell resolution. *GigaScience*, giaa109 (2020).

42. Pan, H., Xue, C., Auerbach, B.J., Fan, J., Bashore, A.C., Cui, J., Yang, D.Y., Trignano, S.B., Liu, W., Shi, J., Ihuegbu, C.O., Bush, E.C., Worley, J., Vlahos, L., Laise, P., Solomon, R.A., Connolly, E.S., Califano, A., **Sims, P.A.**, Zhang, H., Li, M., Reilly, M.P. Single-cell genomics reveals a novel cell state during smooth muscle cell phenotypic switching and potential therapeutic targets for atherosclerosis in mouse and human. *Circulation*, 142, 2060 (2020).

43. Hobson, B.D., Kong, L., Hartwick, E.W., Gonzalez, R.L. Jr., **Sims, P.A.** Elongation inhibitors do not prevent release of puromycylated polypeptide chains from ribosomes. *eLife*, 9, e60048 (2020).

44. Capdevila, C., Calderon, R.I., Bush, E.C., Sheldon-Collins, K., **Sims, P.A.**, Yan, K.S. Single cell transcriptional profiling of the intestinal epithelium. *Methods in Molecular Biology*, 2171, 129 (2020).

45. Mizrak, D., Bayin, N.S., Yuan, J., Liu, Z., Siciu, R., Niphakis, M.J., Ngo, N., Lum, K.M., Cravatt, B.F., Joyner, A.L., **Sims, P.A.** Single-cell profiling and SCOPE-seq reveal the lineage dynamics of adult V-SVZ neurogenesis and NOTUM as a key regulator. *Cell Reports*, 31, 107805 (2020).

46. Martuscello, R.T., Kerridge, C.A., Chatterjee, D., Hartstone, W.G., Kuo, S.H., **Sims, P.A.**, Louis, E.D., Faust, P.L. Gene expression analysis of the cerebellar cortex in essential tremor. *Neuroscience Letters*, 721, 134540 (2020).

47. Szabo, P.A.*, Levitin, H.M.*, Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., Farber, D.L.†, **Sims, P.A.†** Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, 10, 4706 (2019).

48. Sheth, R.U., Li, M., Jiang, W., **Sims, P.A.**, Leong, K.W., Wang, H.H. Spatial metagenomic characterization of microbial biogeography in the gut. *Nature Biotechnology*, 37, 877 (2019).

49. Louis, E.D.†, Kerridge, C.A., Chatterjee, D., Martuscello, R.T., Diaz, D.T., Koeppen, A.H., Kuo, S.H., Vonsattel, J.P.G., **Sims, P.A.**, Faust, P.L.† Contextualizing the pathology in the essential tremor cerebellar cortex: A pathology-omics approach. *Acta Neuropathologica*, 135, 859 (2019).

50. Ma, W., Lee, J., Backenroth, D., Zhou, Y.J., Bush, E., **Sims, P.**, Liu, K., Shen Y., Single cell RNA-Seq reveals pre-cDCs fate determined by transcription factor combinatorial dose. *BMC Molecular and Cell Biology*, 20, 20 (2019).

51. Farber, D.L.†, **Sims, P.A.†** Dissecting lung development and fibrosis at single-cell resolution. *Genome Medicine*, 11, 33 (2019).

52. Hobson, B.D.†, **Sims, P.A.†** Critical analysis of particle detection artifacts in synaptosome flow cytometry. *eNeuro*, 10.1523/ENEURO.0009-19.2019 (2019).

53. Zhang, J., Caruso, F.P., Sa, J.K., Justesen, S., Nam, D.H., **Sims, P.**, Ceccarelli, M., Lasorella, A., Iavarone, A. The combination of neoantigen quality and T lymphocyte infiltrates identifies glioblastomas with the longest survival. *Communications Biology*, 2, 135 (2019).

54. Das Sharma, S.*, Metz, J.B.*, Li, H., Hobson, B.D., Hornstein, N., Sulzer, D., Tang, G., **Sims, P.A.** Widespread alterations in translation elongation in the brain of juvenile *Fmr1* knock-out mice. *Cell Reports*, 26, 3313 (2019).

55. Snyder M.E., Finlayson, M.O., Connors, T.J., Dogra, P., Senda, T., Bush, E., Carpenter, D., Marboe, C., Benvenuto, L., Shah, L., Robbins, H., Hook, J.L., Sykes, M., D'Ovidio, F., Bacchetta, M., Sonett, J.R., Lederer, D.J., Arcasoy, S., **Sims, P.A.**, Farber, D.L. Generation and persistence of human tissue-resident memory T cells in lung transplantation. *Science Immunology*, 4, eaav5581 (2019).

56. Levitin, H.M., Yuan, J., Cheng, Y.L., Ruiz, F.J.R., Bush, E.C., Bruce, J.N., Canoll, P., Iavarone, A., Lasorella, A., Blei, D.M.†, **Sims, P.A.†** *De novo* Gene Signature Identification from Single-Cell RNA-seq with Hierarchical Poisson Factorization. *Molecular Systems Biology*, 15, e8557 (2019).

57. Zhao, J., Chen, A.X., Gartrell, R.B., Silverman, A.M., Aparicio, L., Chu, T., Borbar, D., Shan, D., Samanamud, J., Mahajan, A., Filip, I., Orenbuch, R., Goetz, M., Yamaguchi, J.T., Cloney, M., Horbinski, C., Lukas, R.V., Raizer, J., Rae, A.I., Yuan, J., Canoll, P., Bruce, J.N., Saenger, Y.M., **Sims, P.**, Iwamoto, F.M.†, Sonabend, A.M.†, Rabadan, R.† Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nature Medicine*, 25, 462 (2019).

58. Mizrak, D., Levitin, H.M., Delgado, A.C, Crotet, V., Yuan, J., Chaker, Z., Silva-Vargas, V., **Sims, P.A.†**, Doetsch, F.† Single cell analysis of regional differences in adult V-SVZ neural stem cell lineages. *Cell Reports*, 26, 394 (2019).

59. Yuan, J., Sheng, J., **Sims, P.A.** SCOPE-seq: A Scalable Technology for Linking Live Cell Imaging and Single-Cell RNA-seq. *Genome Biology*, 19, 227 (2018).

60. Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M., Dovas, A., Zanazzi, G., Canoll, P., Bruce, J.N., Lasorella, A., Iavarone, A., **Sims, P.A.** Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Medicine*, 10, 57 (2018).

61. Pan, Y., Duron, C., Bush, E.C., Ma, Y., **Sims, P.A.**, Gutmann D.H., Radunskaya, A., Hardin, J. Graph complexity analysis identifies an ETV5 tumor-specific network in human and murine low-grade glioma. *PLoS One*, 13, e0190001 (2018).

62. Liu, B., Lee, B.W., Nakanishi, K., Villasante, A., Williamson, R., Metz, J., Kim, J., Kanai, M., Bi, L., Brown, K., Di Paolo, G., Homma, S., **Sims, P.A.**, Topkara, V.K., Vunjak-Novakovic, G. Cardiac recovery via extended cell-free delivery of extracellular vesicles secreted by cardiomyocytes derived from iPS cells. *Nature Biomedical Engineering*, 2, 293 (2018).

63. Levitin, H.M., Yuan, J., **Sims, P.A.** Single-cell transcriptomic analysis of tumor heterogeneity. *Trends in Cancer*, 4, 4, 264 (2018).

64. Ding, H., Douglass, Jr., E.F. Sonabend, A.M., Mela, A., Bose, S., Gonzalez, C., Canoll, P., **Sims, P.A.**, Alvarez, M.J., Califano, A., Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nature Communications*, 9, 1471 (2018).

65. Juliano, J., Gil, O., Hawkins-Daarud, A., Noticewala, S., Rockne, R.C., Gallaher, J., Massey, S.C., **Sims, P.A.**, Anderson, A.R.A., Swanson, K.R., Canoll, P. Comparative dynamics of microglial and glioma cell motility at the infiltrative margin of brain tumors. *J.R. Soc. Interface*, 15, 139 (2018).

66. Kiryluk, K., Bomback, A.S., Cheng, Y.L., Xu, K., Camara, P.G., Rabadan, R., **Sims, P.A.**, Barasch, J. Precision medicine for acute kidney injury (AKI): Redefining AKI by agnostic kidney tissue interrogation and genetics. *Seminars in Nephrology*, 1, 40 (2018).

67. Yuan, J., **Sims, P.A.** SPOTS fill a major gap in RNA quantification. *Nature Methods*, 14, 1137-1138 (2017).

68. Bush., E.C., Ray, F., Alvarez, M., Realubit, R., Li, H., Karan, C., Califano, A.†, **Sims, P.A.**† PLATE-Seq for Genome-Wide Regulatory Network Analysis of High-Throughput Screens. *Nature Communications*, 8, 105 (2017).

69. Pan, Y., Bush, E.C., Toonen, J., Ma, Y., Solga, A., **Sims, P.A.**†, Gutmann, D.† Whole tumor RNAsequencing and deconvolution reveal a clinically-prognostic PTEN/PI3K-regulated glioma transcriptional signature. *Oncotarget*, doi: 10.18632/oncotarget.17193 (2017).

70. Kuo, S.H., Lin, C.Y., Wang, J., **Sims, P.A.**, Pan, M.K., Liou, J.Y., Lee, D., Tate, W.J., Kelly, G.C., Louis, E.D., Faust, P.L. Climbing fiber-Purkinje cell synaptic pathology in tremor and cerebellar degenerative diseases, *Acta Neuropathol.,* doi:10.1007/s00401-016-1626-1 (2016).

71. Yuan, J., **Sims, P.A.** An automated microwell platform for large-scale single cell RNA-Seq, *Scientific Reports*, 6, 33883, (2016).

72. Pulko, V., Davies, J.S., Martinez, C., Lanteri, M.C., Busch, M.P., Diamond, M.S., Knox, K., Bush, E.C., **Sims, P.A.**, Sinari, S., Billheimer, D., Haddad, E.K., Murray, K.O., Wertheimer, A.M., Nikolich-Zugich, J. Human memory T cells with a naive phenotype accumulate with aging and respond to persistent viruses, *Nature Immunology*, 17, 966, (2016).

73. Sims, J.S., Grinshpun, B., Feng, Y., Ung, T.H., Niera, J.A., Samanamud, J.L., Canoll, P., Shen, Y.†, **Sims, P.A.**†, Bruce, J.N.† Diversity and Divergence of the glioma-infiltrating T cell receptor repertoire, *Proc. Natl. Acad. Sci. USA*, doi:10.1073/pnas.1601012113 (2016).

74. Hornstein, N.*, Torres, D.*, Das Sharma, S., Tang, G., Canoll, P.†, **Sims, P.A.**† Ligation-free ribosome profiling of cell type-specific translation in the brain, *Genome Biology*, 17, 149 (2016).

75. Lescarbeau, R., Lei, L., Bakken, K.K., **Sims, P.A.**, Sarkaria, J.N., Canoll, P., White, F.M. Quantitative phosphoproteomics reveals Wee1 kinase as a therapeutic target in a model of proneural glioblastoma, *Molecular Cancer Therapeutics*, 15, 1332 (2016).

76. Bose, S., **Sims, P.A.** Genome-Wide Analysis of Single Cells and the Role of Microfluidics. *Wiley Blackwell Biotechnology Series: Micro and Nanosystems for Biotechnology*, (2016).

77. Qiu, W., Tang, S.M., Lee, S., Turk, A.T., Sireci, A.N., Qiu, A., Rose, C., Xie, C., Kitajewski, J., Wen, H., Crawford, H.C., **Sims, P.A.**, Hruban, R.H., Remotti, H.E., Su, G. Loss of Activin Receptor Type 1B Accelerates Development of Intraductal Papillary Mucinous Neoplasms in Mice with Activated KRAS. *Gastroenterology*, 150, 218-228 (2016).

78. Venere, M., Horbinski, C., Crish, J.F., Jin, X., Vasanji, A., Major, J., Burrows, A.C., Chang, C., Prokop, J., Wu, Q., **Sims, P.A.**, Canoll, P., Summers, M.K., Rosenfeld, S.S., Rich, J.N. The mitotic kinesin KIF11 is a driver of invasion, proliferation, and self-renewal in glioblastoma. *Science Translational Medicine*, 7, 304ra143 (2015).

79. Bose, S., Wan, Z., Carr, A., Rizvi, A.H., Vieira, G., Pe'er, D., **Sims, P.A.** Scalable Microfluidics for Single Cell RNA Printing and Sequencing. *Genome Biology*, 16, 120 (2015).

80. Shelanski, M., Shin, W., Aubry, S., **Sims, P.**, Alvarez, M.J., Califano, A. A Systems Approach to Drug Discovery in Alzheimer's Disease. *Neurotherapeutics*, 12, 126-131 (2015).

81. Worthley, D.L., Churchill, M., Compton, J.T., Tailor, Y., Rao, M., Si, Y., Levin, D., Schwartz, M.G., Uygur, A., Hayakawa, Y., Gross, S., Renz B.W., Setlik, W., Martinez, A.N., Chen, X., Nizami, S., Lee, H.G., Kang, H.P., Caldwell, J.M., Asfaha, S., Westphalen, C.B., Graham, T., Jin, G., Nagar, K., Wang, H., Kheirbek, M.A., Kolhe, A., Carpenter, J., Glaire, M., Nair, A., Renders, S., Manieri, N., Muthupalani, S., Fox, J.G., Reichert, M., Giraud, A.S., Schwabe, R.F., Pradere, J.P., Walton, K., Prakash, A., Gumucio, D., Rustgi, A.K., Stappenbeck, T.S., Friedman, R.A., Gershon, M.D., **Sims, P.**, Grikscheit, T., Lee, F.Y., Karsenty, G., Mukherjee, S., Wang, T.C. Gremlin 1 Identifies a Skeletal Stem Cell with Bone, Cartilage, and Reticular Stromal Potential. *Cell*, 160, 269-284 (2015).

82. Gill, B.*, Pisapia, D.*, Malone, H., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J., Banu, M., Dovas, A., Teich, A.F., Sheth, S., McKhann, G., Sisti, M., Bruce, J.N.†, **Sims, P.A.†** , Canoll, P.† MRI-Localized Biopsies Reveal Subtype-Specific Differences in Molecular and Cellular Composition at the Margins of Glioblastoma. *Proc. Natl. Acad. Sci. USA*, 111, 12550-12555 (2014).

83. Gonzalez, C., Sims, J.S.*, Hornstein, N.*, Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P.†, **Sims. P.A.†** Ribosome Profiling Reveals a Cell Type-Specific Translational Landscape in Brain Tumors. *J. Neuroscience*, 34, 10924-10936 (2014).

84. Sonabend, A.M., Bansal, M., Guarnieri, P., Lei, L., Amendolara, B., Soderquist, C., Leung, L., Yun, J., Kennedy, B., Sisti, J., Bruce, S., Bruce, R., Shakya, R., Ludwig, T., Rosenfeld, S., **Sims, P.A.**, Bruce, J.N., Califano, A., Canoll, P. The Transcriptional Regulatory Network of Proneural Glioma Determines the Genetic Alterations Selected during Tumor Progression. *Cancer Research*, 74, 1440 (2014).

85. Sonabend, A.M., Yun, J., Lei, L., Leung, R., Soderquist, C., Crisman, C., Gill, B.J., Carminucci, A., Sisti, J., Castelli, M., **Sims, P.A.**, Bruce, J.N., Canoll, P. Murine cell line model of proneural glioma for evaluation of anti-tumor therapies. *J. Neuro-Oncology*, 112, 375 (2013).

86. Lu, P.J., Hoehl, M.M., Macarthur, J.B., **Sims, P.A.**, Ma, H., Slocum, A.H. Robust and economical multi-sample, multi-wavelength UV/vis absorption and fluorescence detector for biological and chemical contamination. *AIP Advances*, 2, 032110 (2012).

87. Hoehl, M.M., Lu, P.J., **Sims, P.A.**, Slocum, A.H. Rapid and Robust Detection Methods for Poison and Microbial Contamination. *J. Agric. Food Chem*., 60, 6349-6358 (2012).

88. Men, Y., Fu, Y., Chen, Z., **Sims, P.A.**, Greenleaf, W.J., Huang, Y. Digital Polymerase Chain Reaction in an Array of Femtoliter Polydimethylsiloxane Microreactors. *Analytical Chemistry*, 84, 4262-4266 (2012).

89. Shiroguchi, K., Jia, T.Z., **Sims, P.A.**, Xie X.S. Digital RNA Sequencing Minimizes SequenceDependent Bias and Amplification Noise with Single Molecule Barcodes. *Proc. Natl. Acad. Sci. USA*, 109, 1347-1352 (2012).

90. **Sims, P.A.***, Greenleaf, W.J.*, Duan, H., Xie, X.S. Fluorogenic DNA Sequencing in PDMS Microreactors. Nature Methods, 8, 575-580 (2011).

91. **Sims, P.A.**, Xie, X.S. Probing dynein and kinesin stepping with mechanical manipulation in a living cell. *ChemPhysChem*, 10, 1511-1516 (2009).

92. Sims, J.S., Militello, K.T., **Sims, P.A.**, Patel, V.P., Kasper, J.M., Wirth D.F. Patterns of genespecific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle. *Eukaryot. Cell.* 8, 327-338 (2009).

93. Nan, X., **Sims, P.A.**, Xie, X.S. Organelle tracking in a living cell with microsecond time resolution and nanometer spatial precision. *ChemPhysChem*, 9, 707-712 (2008).

94. Lu, P.J., **Sims, P.A.**, Oki, H., MacArthur, J., Weitz, D.A. Target-locking acquisition with real-time confocal (TARC) microscopy. *Optics Express*, 15, 8702-8712 (2007).

95. Nan, X., **Sims, P.A.**, Chen, P., Xie, X.S. Observation of individual microtubule motor steps in living cells with endocytosed quantum dots. *J. Phys. Chem. B*, 109, 24220-24224 (2005).

96. **Sims, P.A.**, Wong, C.F., Vuga, D., McCammon, J.A., Sefton, B.M. Relative contributions of desolvation, inter-, and intramolecular interactions to binding affinity in protein kinase systems. *J. Comp. Chem*., 26, 668-681 (2005).

97. **Sims, P.A.**, Wong, C.F., McCammon, J.A. Charge optimization of the interface between protein kinases and their ligands. *J. Comp. Chem.*, 25, 1416-1429 (2004).

98. **Sims, P.A.**, Wong, C.F., McCammon, J.A. A computational model of binding thermodynamics: The design of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem*., 46, 3314-3325 (2003).

Awarded Patents

1. **Sims, P.A.** High-Throughput Single Molecule Protein Identification. U.S. Patent 11,927,593 (2024).

2. Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,866,781 (2024).

3. Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,834,712 (2023).

4. **Sims, P.A.**, Bose, S., Yuan, J. RNA Printing and Sequencing Devices, Methods, and Systems. U.S. Patent 11,788,120 (2023).

5. Xie X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,629,379 (2023).

6. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,608,527 (2023).

7. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,352,669 (2022).

8. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,286,523 (2022).

9. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,078,533 (2021).

10. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 11,035,001 (2021).

11. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 10,584,382 (2020).

12. **Sims, P.A.** High-Throughput Single Molecule Protein Identification. U.S. Patent 10,466,248 (2019).

13. Xie, X.S., Shiroguchi, K., **Sims, P.A.**, Jia, T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 10,287,630 (2019).

14. Xie X.S., Shiroguchi K., **Sims P.A.**, Jia T.Z. Single Cell Nucleic Acid Detection and Analysis. U.S. Patent 9,260,753 (2016).

Invited Presentations

Single-Cell Analysis of Early and Memory Responses to Vaccines in Human Tissues, Systems Immunology in Aging and Complex Diseases Meeting, The Jackson Laboratory, Farmington, CT, 2024.

Systems Immunology and Machine Learning Applications Across Human Tissues, Artificial Intelligence and Immunology Meeting, National Institute of Allergy and Infectious Disease, Rockville, MD, 2024.

New Tools for Multi-Modal Single-Cell Analysis, Single Cell Analysis Boot Camp Keynote Lecture, Mailman School of Public Health, Columbia University, NY, 2024.

Cell Type-Specific Drug Responses in the Glioma Microenvironment, Tumor-Immune Systems Biology Symposium, Memorial Sloan Kettering Cancer Center, NY, 2024.

New Tools for Multi-Modal Single-Cell Analysis, Single Cell Analysis Boot Camp Keynote Lecture, Mailman School of Public Health, Columbia University, NY, 2023.

New Tools for Multi-Modal Single-Cell Analysis, Keynote Lecture, Single Cell Symposium, Johns Hopkins University, Baltimore, MD, 2023.

New Tools for Multi-Modal Single-Cell Analysis in Cancer and Immunology, Keynote Lecture, Berrie Diabetes Center Scientific Retreat, 2023.

Scalable Co-Sequencing of RNA and DNA from Individual Nuclei with DEFND-seq, 10x Genomics User Group Meeting, Philadelphia, PA, 2023.

New Tools for Multi-Modal Single-Cell Analysis, Single Cell Analysis Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2023.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, 23andMe Therapeutics (virtual), 2023.

Cell Type-Specific Drug Responses in Patient-Derived Glioma Slices, The 2nd Wyss Family Think Tank on Genetic Predisposition to Primary CNS Cancer, Pismo Beach, CA, 2023.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Single Cell Analysis Boot Camp Keynote Lecture, Mailman School of Public Health, Columbia University, NY, 2023.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Wellcome Sanger Institute, Hinxton, UK, 2023.

Single-Cell Genomics and Why Flow Cytometry is More Important than Ever, MetroFlow Annual Meeting, Union, NJ, 2022.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Single Cell Analysis Boot Camp Keynote Lecture, Mailman School of Public Health, Columbia University, NY, 2022.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Single Cell Analysis Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2022.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, AstraZeneca Oncology, 2022.

SCOPE-seq: Scalable Technology for Linking Live Cell Imaging and scRNA-seq, Biophysical Society Japan, 2021.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Keynote Lecture, Radiation Oncology / Center for Radiation Research Retreat, Columbia University, New York, NY, 2021.

Subcellular Molecular Profiling of Midbrain Dopamine Neurons, Cajal Neuroscience, 2021.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Magee-Womens Research Institute, University of Pittsburgh, 2021

An Integrated Map of the Immune System in Humans Across Tissues and Ages, Chan Zuckerberg Initiative Seed Networks Annual Meeting (virtual), 2020

scRNA-seq and SCOPE-seq Uncover a Secreted Enzyme Regulator of Adult Neurogenesis in the Mouse Brain, American Society for Cell Biology (ASCB) / European Molecular Biology Organization (EMBO) Joint Meeting, Washington D.C., 2019

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, International Symposium of the Princess Takamatsu Cancer Research Fund, Tokyo, Japan, 2019.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Physical Sciences Oncology Center Seminar, University of Pennsylvania, Philadelphia, PA 2019.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Institute for Systems Genetics Seminar, New York University, New York, NY 2019.

Single-Cell RNA-seq Uncovers a Secreted Enzyme Regulator of Adult Neurogenesis in the Mouse Brain, Single Cell Analysis Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 2019.

Single-Cell Analysis of Differentiation and Drug Response in Glioblastoma, Pathology Grand Rounds, Renaissance School of Medicine, Stony Brook University, Stony Brook, NY 2019.

A Single-Cell Reference Map for Human Blood and Tissue T Cell Activation, 10x User Group Meeting, New York, NY 2019.

Single Cell Transcriptomics, The Genome Access Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2019.

New Tools for Single-Cell Analysis in Cancer: Differentiation and Drug Response in Glioblastoma, Cardiovascular Seminar Series, Columbia University, New York, NY, 2019.

Differentiation and Drug Response in Glioblastoma, Experimental Biology: Histochemical Society Workshop, Orlando, FL, 2019.

Deconvolving Cell Type-Specific Drug Response in Brain Tumors with Single-Cell Genomics, New York Genome Center, New York, NY, 2019.

High-Dimensional Single Cell Analysis of Lineage Diversity and Drug Response in Glioblastoma, Grand Rounds, Department of Neurological Surgery, Columbia University, New York, NY, 2019.

Discussion Leader: Precision Medicine and Cancer Evolution in Clinical Trial Design, Gordon Research Conference: Physical Science of Cancer, Galveston, TX, 2019.

SCOPE-seq: A Scalable Technology for Linking Live Cell Imaging and Single-Cell RNA-seq, IMAT Investigators Meeting, National Cancer Institute, Rockville, MD, 2018.

High-Dimensional Single Cell Analysis of Glioblastoma, Neuro-oncology Program, Herbert Irving Comprehensive Cancer Care Center, Columbia University, New York, NY, 2018

New Tools for Single-Cell Analysis in Cancer, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 2018.

High-Dimensional Single-Cell Analysis of Glioblastoma, Department of Physiology and Biophysics, College of Medicine, University of Illinois, Chicago, IL 2018.

Probing Translational Dysregulation in Neurological Disorders with High-Sensitivity Ribosome Profiling, Royal Society, Chicheley Hall, England 2018.

Single-Cell Dissection of High-Grade Glioma. Cancer Genomics and Mathematical Data Analysis Symposium, Columbia University Medical Center, New York, NY 2018.

Large-Scale Integration of Live Cell Imaging and Single-Cell RNA-Seq. IMAT Investigators Meeting, National Cancer Institute, Rockville, MD, 2017.

Single-Cell Dissection of High-Grade Glioma. Department of Neurology, Yale School of Medicine, New Haven, CT, 2017.

Single-Cell Dissection and Topological Analysis of High-Grade Glioma. Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA, 2017.

Applications and Analysis of Large-Scale, Single-Cell RNA-Seq. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2017.

New Tools for Single Cell Analysis in Cancer. NCI Special Session on Technologies for Advanced Omic Analysis, American Association for Cancer Research (AACR) Annual Meeting, Washington D.C., 2017

High-Dimensional Single Cell Analysis of Glioblastoma: New Tools and Insights. O'Brien Urology Symposium, Columbia University, New York, NY, 2017.

High-Dimensional Single Cell Analysis of Glioblastoma: New Tools and Insights. Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 2017.

High-Dimensional Single Cell Analysis of Glioblastoma: New Tools and Insights. Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, 2017.

High-Dimensional Single Cell Analysis of Glioblastoma. Cancer Regulatory Networks Seminar, Columbia University Medical Center, New York, NY, 2017.

Microfluidics for Large-Scale Single Cell RNA-Seq in Glioblastoma. Molecular Medicine Tri-Conference, San Francisco, CA, 2017.

Scalable Microfluidics for Linking Single Cell Microscopy and Sequencing. Single Cell Analysis Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2016.

Single Cell and Cell Type-Specific Expression Analysis in Neurological Disorders. Epilepsy Research Seminar, Department of Neurology, Columbia University Medical Center, New York, NY 2016.

Scalable Microfluidics for Linking Single Cell Microscopy and Sequencing. The Center for Computational Biology, The University of Kansas, Lawrence, KS, 2016.

Scalable Microfluidics for Linking Single Cell Microscopy and Sequencing. Five Points Lecture, New York Genome Center, New York, NY, 2016.

A Scalable Platform for Single Cell Transcriptomics. Gordon Research Conference: Molecular Mechanisms in Evolution, Easton, MA, 2015.

A Scalable Platform for Single Cell RNA-Seq. Clontech Laboratories, Mountain View, CA, 2015.

A Scalable Platform for Single Cell RNA-Seq. Single Cells Meeting, Cold Spring Harbor Asia, Suzhuo, China, 2014.

Microwell Arrays for Single Cell Transcriptomics. Single Cell Analysis Course, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2014.

Image-Guided RNA-Seq Reveals Subtype-Specific Patterns at the Infiltrative Margins of Glioblastoma. Tumor Microenvironment-New York City (TME-NYC) Symposium, NYU School of Medicine and Columbia University Medical Center, New York, NY, 2014.

Tackling Cellular Heterogeneity in Glioma. Grand Rounds, Department of Neurological Surgery, Columbia University Medical Center, New York, NY, 2013.

Tackling Cellular Heterogeneity in Systems Biology. MIT Biophysics Program, Massachusetts Institute of Technology, Cambridge, MA, 2013.

Microreactor Arrays for Digital PCR, Sequencing, and Single Cell Analysis. Biochemistry and Biotechnology Program, University of Missouri at St. Louis, St. Louis, MO, 2012.

Microreactor Arrays for Digital PCR, Sequencing, and Single Cell Analysis. Digital PCR Application and Advances Conference, Cambridge Healthtech Institute, San Diego, CA, 2012.

Fluorogenic Pyrosequencing. Columbia Initiative in Systems Biology, College of Physicians and Surgeons, Columbia University, New York, NY, 2011.

Fluorogenic Pyrosequencing in PDMS Microreactors. Biodynamic Optical Imaging Center, School of Life Sciences, Peking University, Beijing, China, 2011.

Optical trapping of kinesin- and dynein-driven cargoes in living cells with submillisecond tracking. Single Molecule Discussion Group, Harvard Medical School, Boston, MA, 2008.

Professional Service

Editorial

Editorial Board, *Genome Medicine*, 2022-present

Guest Editor, *Current Opinion in Systems Biology*, Genetics and Epigenomics, 2018

Grant Review

NIH Review Panel on Biomaterials, Drug Delivery, and Instrumentation R01 Special Emphasis Panel, 2024

NIH Small Business: Cell and Molecular Biology Panel, 2023

NIH Common Fund, NIH Director's Transformative Research Award TR01 Special Emphasis Panel, 2023

NIH National Institute of Diabetes and Digestive and Kidney Diseases RC2 Special Emphasis Panel, 2022

NIH Human Biomolecular Atlas Program (HuBMAP) U54 Special Emphasis Panel, 2022

The Mark Foundation for Cancer Research, Aspire Award Review, 2021

NIH National Institute of Mental Health UG3 Special Emphasis Panel: BRAIN Initiative, 2021

NIH National Institute of Mental Health R01 Special Emphasis Panel: BRAIN Initiative, 2021

NIH National Institute of Diabetes and Digestive and Kidney Diseases, Consortium of Beta Cell Death and Survival U01 Special Emphasis Panel, 2020

NIH National Institute of Mental Health, BRAIN Initiative: Novel Tools for Cells and Circuits R01/UG3/UH3 Special Emphasis Panel, 2020

The Mark Foundation for Cancer Research, Aspire Award Review, 2019

NIH National Institute of Neurological Disorders and Stroke P01 Special Emphasis Panel, 2019

The Emerson Collective Cancer Research Fund, 2019.

NIH Office of the Director DP5 Special Emphasis Panel for Early Independence Award, 2018.

NIH IMST Small Business: Cell and Molecular Biology Panel, 2018.

NIH National Institute of Mental Health R01 Special Emphasis Panel: BRAIN Initiative, 2018

NIH National Cancer Institute U01 Special Emphasis Panel: Cancer Systems Biology Consortium, 2017

NIH National Institute of Mental Health U01 Special Emphasis Panel: BRAIN Initiative, 2017

NIH National Cancer Institute U01 Special Emphasis Panel: Physical Sciences Oncology Projects, 2017

NIH National Institute of Mental Health U01 Special Emphasis Panel: BRAIN Initiative, 2016

NIH National Institute of Allergy and Infectious Disease P01 Special Emphasis Panel, 2016

NIH National Institute of Mental Health R01 Special Emphasis Panel: Clinical Studies of Mental Illness Not Involving Treatment Development, Efficacy, or Effectiveness Trials, 2016

Swiss National Science Foundation Research Equipment Program Panel, 2015

NIH National Institute of Mental Health R01 Special Emphasis Panel: Clinical Studies of Mental Illness Not Involving Treatment Development, Efficacy, or Effectiveness Trials, 2015

NIH National Institute of Mental Health U01 Special Emphasis Panel: BRAIN Initiative, 2014

NIH National Institute of Mental Health R01 Special Emphasis Panel: EUREKA for Neuroscience and Disorders of the Nervous System, 2014

NIH National Institute of Mental Health R01 Special Emphasis Panel: EUREKA for Neuroscience and Disorders of the Nervous System, 2013

NIH National Institute of Mental Health R21 Special Emphasis Panel: Development of New Tools to Explore the Synaptome, 2012

Journal Review

Biophysical Journal, Cancer Research, Cell, Cell Genomics, Cell Reports, Cell Reports Medicine, Cell Reports Methods, CellStemCell, Cell Systems, ChemPhysChem, Communications Biology, eLife, Frontiers in Immunology, Genome Biology, Genome Medicine, Genome Research, Immunity, iScience, Journal of the Royal Society Interface, Journal of Translational Medicine, Lab on a Chip, Molecular Cancer Research, Molecular Systems Biology, Nature, Nature Biomedical Engineering, Nature Biotechnology, Nature Cancer, Nature Communications, Nature Genetics, Nature Immunology, Nature Medicine, Nature Methods, Nature Neuroscience, Nature Reviews Genetics, Neuron, Nucleic Acids Research, PeerJ, PNAS, Science, Science Advances, Science Immunology, Trends in Cancer

Teaching and Mentoring

Courses

Lecturer in Genetics Approaches II, 2023
Lecturer in Cellular and Molecular Biology of Cancer, 2018-2022
Lecturer in Biochemistry, Molecular & Cell Biology I, 2013-2021
Instructor and Lecturer in Deep Sequencing, 2015-2019; 2021-2023
Lecturer in MD/PhD Ethics in Research, 2014-2023
Lecturer in Columbia Med into Grad Program, 2017
Lecturer in Statistics for Basic Sciences, 2018-2022
Instructor in Single Cell Analysis Course, Cold Spring Harbor Laboratory, 2014-2019, 2022-2023

Graduate Student Researchers

Stephanie Andersen (MD/PhD Program)
Romella Sagatelian (Systems Biology PhD Program)
Shannon Weng (MD/PhD Program)
Abby Brand (Integrated PhD Program)
Jingying Huang (Biotechnology MA Program)
Kwasi Osae-Kwapong (NYU Bioinformatics MA Program)
Alex George (MD/PhD Program)
Daniel Caron (Microbiology, Immunology, and Infection PhD Program)
Qiuyu Gong (Biotechnology MA Program)
John Tuddenham (MD/PhD Program, defended 12/2022)
Zhouzerui Liu (Systems Biology PhD Program, defended 11/2023)
Benjamin Hobson (MD/PhD Program, defended 9/2021)
Hanna Mendes Levitin (Systems Biology PhD Program, defended 4/2020)
Jordan Metz (MD/PhD Program, defended 12/2021)
Nathan Omans (Neuroscience MS Program)
Nicholas Hornstein (MD/PhD Program, defended 12/2016)
Erin Bush (Biotechnology MA Program)

Postdoctoral Research Scientists

Josue Curto-Navarro, PhD
Ana Paula Delgado, PhD
Timothy Olsen, PhD
Katherine Xu, PhD
Wenting Zhao, PhD
Yim Ling Cheng, PhD
Dogukan Mizrak, PhD
Jinzhou Yuan, PhD
Sohani Das Sharma, PhD
Gregory Vieira, PhD
Christian Gonzalez, PhD
Sayantan Bose, PhD