

Case 2024-01419
U.S. Patent No. 11,840,730

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

**ILLUMINA INC.,
Petitioner**

v.

**MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner**

Case IPR2024-01419

U.S. Patent No. 11,840,730

**TITLE: Methods and Compositions For Evaluating Genetic Markers
Issue Date: December 12, 2023**

DECLARATION OF JOSEPH D. PUGLISI, PH.D.

IPR2024-01419

Illumina Inc., v. Molecular Loop Biosciences, Inc.

Molecular Loop's Exhibit 2023

TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION	5
II. BASES FOR OPINIONS	5
III. MATERIALS REVIEWED	6
IV. EDUCATION AND EXPERIENCE.....	6
V. LEGAL STANDARDS	10
A. Anticipation	10
B. Obviousness.....	11
VI. LEVEL OF ORDINARY SKILL IN THE ART	13
VII. TECHNICAL BACKGROUND	15
A. Sanger-Style Sequencing.....	15
B. Next Gen Sequencing (NGS)	20
C. Synthetic Biology	24
D. Genotyping	28
VIII. The '730 Patent.....	28
A. Background	28
B. The Invention Of The '730 Patent Addressed Accuracy Issues With Next-Gen Sequencers	30
C. Prosecution History of the '730 Patent	35
D. The '730 Patent Is Entitled To A Priority Date Of At Least April 29, 2009	37
E. A POSA Would Have Understood That The March 6, 2009 Documents Discloses Claims 1-3 and 9 Elements Of The '730 Patent	38

1.	Claim 1	38
2.	Claim 2	44
3.	Claim 3	46
4.	Claim 9	47
F.	A POSA Would have Understood That The May 18, 2009 Provisional Application Discloses Claim 1-3 and 9 Of The '730 Patent.	70
1.	Claim 1	71
2.	Claim 2	77
3.	Claim 3	78
4.	Claim 9	80
IX.	ILLUMINA'S ASSERTED PRIOR ART	114
A.	Chee	114
B.	Gahan	121
C.	McCloskey	122
D.	Bignell	125
E.	Chiang	127
F.	Bentley	128
X.	GROUND 1-2: THE CHALLENGED CLAIMS ARE NOT DISCLOSED BY CHEE	130
A.	Petitioner And Dr. Gabriel Fail To Show That Chee Discloses A Method For "collapsing target:differentiator tag combinations observed more than once into a single count."	130
1.	The '730 Patent Discloses The "Collapsing" Step As A Limitation Separate And Distinct From "Determining A Consensus Base Call."	131

2.	Chee Does Not Disclose The Collapsing Step.	140
i.	Chee’s Generic Disclosures About Skewing And Amplification Bias Do Not Disclose A Method For “Collapsing” As Required By The ’730 Patent.....	142
ii.	Chee’s Disclosures Show Reading Multiple Clonal Copies Of The Same Molecule, Which Is The Opposite Of “Collapsing . . . Into A Single Count.”	147
B.	Petitioner And Dr. Gabriel Fail To Show That Chee’s Provisional Application Discloses A Method For “obtaining nucleic acids from a biological sample.”	152
XI.	GROUND 3: THE CHALLENGED CLAIMS ARE PATENABLE OVER CHEE IN VIEW OF GAHAN.....	155
XII.	GROUND 4: THE CHALLENGED CLAIMS ARE PATENTABLE OVER MCCLOSKEY IN VIEW OF BIGNELL.....	161
A.	McCloskey And Bignell Do Not Disclose Or Render Obvious The Limitations of Claim 1	161
B.	A POSA Would Not Have Been Motivated To Combine Bignell and McCloskey.....	167
1.	McCloskey And Bignell Teach Away From Each Other.	167
2.	It Would Have Been Cost-Prohibitive To Combine McCloskey And Bignell Because They Are Directed To Different Sequencing Technologies.....	168
XIII.	GROUND 5: THE CHALLENGED CLAIMS ARE PATENTABLE OVER MCLLOSKEY IN VIEW OF BIGNELL AND BENTLEY.....	174
XIV.	GROUND 6: THE CHALLENGED CLAIMS ARE PATENTABLE OVER MCCLOSKEY IN VIEW OF BIGNELL AND GAHAN	175
XV.	GROUND 7: THE CHALLENGED CLAIMS ARE PATENTABLE OVER MCCLOSEY IN VIEW OF BIGNELL AND CHIANG	175
XVI.	APPENDIX A: THE CHALLENGED CLAIMS OF THE ’730 PATENT	177

I. INTRODUCTION

1. I, Joseph Daniel Puglisi, Ph.D., a resident of 773 Mayfield Ave., Stanford, California, over 18 years of age, hereby declare as follows:
 - I have personal knowledge of all of the matters about which I testify in this declaration.
 - I have been retained on behalf of Patent Owner Molecular Loop to provide an independent expert opinion and testimony about U.S. Patent No. 11,840,730 (the “’730 patent”) and its Claims 1–9 (the “Challenged Claims”). The full text of the Challenged Claims appears in Appendix A to this declaration.
 - I am being compensated for my time spent consulting on this matter at my normal hourly rate and I am receiving reimbursement for expenses incurred in the course of my work. My compensation does not depend in any way on either the opinions I have reached or the outcome of this proceeding.

II. BASES FOR OPINIONS

2. I have considered and reviewed the documents and materials listed below in Section III in light of my specialized knowledge provided by my education, training, research, and experience, as summarized in Section IV and described in my CV attached to my declaration. My analysis of

the documents and materials in combination with the specialized knowledge that I have obtained over the course of my education and career, form the bases for my opinion.

III. MATERIALS REVIEWED

3. In forming the opinions set forth in this declaration, I have relied upon:
- my knowledge and experience in the field of sequencing technologies, genetics, advanced genomic technologies, molecular biology, biochemistry, and diagnostic testing;
 - U.S. Patent No. 11,840,730;
 - the prosecution history of U.S. Patent No. 11,840,730;
 - Illumina's Petition for *inter partes* review and the materials cited therein;
 - The Board's Institution Decision and the materials cited therein;
 - the declaration of Dr. Stacey Gabriel (EX1005) and the materials cited therein as well as the transcript of the July 9, 2025 deposition testimony of Dr. Gabriel (EX2024); and
 - all other materials cited herein.

IV. EDUCATION AND EXPERIENCE

4. I am currently the Jauch Professor and Professor of Structural Biology at the Stanford University School of Medicine. I have 40 years of

experience in the investigation of nucleic acid and protein chemistry, biophysics, and biology as a professor, consultant and as a student.

5. I am an expert in biophysics, biochemistry, molecular biology, cell diagnostics, molecular interactions, DNA and RNA sequencing, genetic testing, nucleic acid preservation, and related fields. My qualifications and credentials are set forth in my curriculum vitae, attached. I received a B.A. in Chemistry in 1984 from The Johns Hopkins University and a Ph.D. in Biophysical Chemistry in 1989 from the University of California, Berkeley. While pursuing my Ph.D., I received the Bruce Mahan Memorial Teaching Award from the University of California Berkeley's Department of Chemistry.
6. From 2004 through 2014, I was the Chair of Stanford's Department of Structural Biology. I was the Director of the Stanford Magnetic Resonance Laboratory from 1997-2015. I have also been the Director of the International School of Biological Magnetic Resonance since 2003. From 1993 through 1997, I was part of the faculty at the University of California, Santa Cruz in the Molecular, Cellular, & Developmental Biology Department. From 1991 to 1993, I was a postdoctoral fellow at the Massachusetts Institute of Technology in the Department of Chemistry. From 1989 to 1991, I was a European Molecular Biology

Organization (EMBO) postdoctoral fellow at the Institute of Molecular and Cellular Biology at the French National Centre for Scientific Research.

7. I have authored or co-authored over 175 publications in the fields of biophysics, chemistry, molecular and cell biology, including at least five book chapters. I have co-authored a textbook on Physical Chemistry. I have been invited to present on topics in the fields of molecular and cell biology on at least 300 occasions, and presented numerous named Lectureships, including the Linus Pauling Lecture (Stanford), Tom Steitz Lecture (Yale), Ferry Lecture (University of Wisconsin), O'Keefe Lecture (Arizona State University).
8. I am a member of numerous technical societies, including the Biophysical Society and American Chemical Society. I was a council member of the Biophysical Society from 2014–2017. I have been the lead organizer of the past 12 NATO Advanced Study Institute courses on the fields of structural biology and molecular biophysics. In 2007, I was the co-chair for both the Nucleic Acids Gordon Conference hosted by the Salve Regina University, and the American Chemical Society's RNA Biophysics Symposium.
9. I have been appointed as an editor of a number of scientific publications.

I have been a member of the Editorial Board for the Proceedings of the National Academies of Science since 2017. I have been an Editor of the Biophysical Journal since 2014. I have been a member of the Advisory Editorial Board for EMBO reports since 2011. I have been an editor of the Structure division of the prestigious Cell Press since 2007, and previously served as the Senior Editor from 2003-2007.

10. In 2014, I was elected to the US National Academies of Science. I received the Merit Award from the National Institute of Health in 2011. I also received the National Institute of Health Director's Transformative R01 Program Award in 2011. I was an Alfred P. Sloan Research Fellow for the Alfred P. Sloan Research Foundation in 1997. I was also the David and Lucille Packard Fellow for the David and Lucille Packard Fellowship in Science and Engineering from 1994–1999. I thrice received the University of California at Berkeley's Bruce Mahan Memorial Teaching Award in the Department of Chemistry. In 1989, I was inducted into the Sigma Xi Honor Society.
11. My research for the past four decades involved the study of nucleic acid structure and function. My research includes the investigation of RNA function by studying molecular structure and dynamics using a variety of biophysical and biological tools, such as nuclear magnetic resonance

(NMR) spectroscopy, x-ray crystallography, cryogenic electron microscopy, and single-molecule fluorescence methods. I have developed and characterized methods to synthesize, purify, characterize, and stabilize highly pure and structurally stable RNA molecules for these applications, and applied/developed DNA/RNA analysis tools, including sequencing methods.

V. LEGAL STANDARDS

12. Molecular Loop's counsel has helped me understand the applicable legal standards for this case. I am not an attorney and do not have any formal training in patent law. My opinions are based on my understanding of the legal concepts as I describe them below.

A. Anticipation

13. I understand that a patent may be anticipated and invalid only when each and every element of a patent is found in a single prior art reference arranged as in the claim. I understand that, for anticipation purposes, two references cannot be combined (even if one reference incorporates another by reference) unless the allegedly anticipatory reference has a particularized identification of the material it incorporates and a clear indication of where the incorporated material may be found in the second reference.

14. I further understand that prior art publications must enable a POSA to make the invention without undue experimentation. To anticipate a patent, I understand that each disclosed feature must either be necessarily present in the prior art reference or the natural result of explicitly disclosed elements in the prior art reference.

B. Obviousness

15. I understand a claim of a patent will be found invalid and obvious only if each element of the claim would have been obvious to a POSA at the time of the invention when comparing the claim to the prior art references. I further understand that a claim will be obvious if more than one prior art reference disclose every claim element such that it would have made the invention obvious to a POSA at the time of the invention.
16. I understand that determining whether an invention is obvious involves making decisions on fact issues, but the ultimate obvious determination is a question of law. I understand relevant factual questions include: (1) determining the level of skill that a POSA would have at the time of the invention, (2) determining the scope and content of the prior art, (3) considering how the claimed invention as a whole is different from the prior art, and (4) considering objective secondary considerations that reflect the contemporaneous response to the invention at the time, such

as a failure of others, commercial success, and long-felt need.

17. I understand that the obviousness analysis requires considering the state of the art at the time of the invention. I took care to avoid importing hindsight bias into my obviousness analysis. I also understand that using the invention as a model for putting together pieces of prior art to form the invention is improper. For example, if a POSA at the time of the invention knew of the problem and was instructed to make the patented invention, the POSA would almost always be able to make the invention. Thus, in my obviousness analysis I considered whether the inventive concept included discovering a problem existed or defining the problem in a new way. I understand that an obviousness determination requires a showing by a preponderance of evidence that a POSA would have recognized the problem identified by the inventor and found it obvious to use the inventor's methods as a solution to that problem.
18. I understand that the burden to prove obviousness lies on the party challenging the patent. I understand that the obviousness analysis requires showing that a POSA at the time of the invention would have had a motivated to combine the elements in the manner claimed by the patent with a reasonable expectation of success. While a prior art reference may expressly or inherently disclose a claim limitation, I

understand that a claim limitation is not proven obvious just by showing that each of the elements were independently known in the prior art. Instead, I understand that there must be a reason why it would have been obvious to change the prior art to arrive at the invention.

19. I understand that conclusory statements are insufficient to show obviousness and I have considered whether Dr. Gabriel and Petitioner have provided support to explain why they arrive at their conclusion that the invention is obvious.
20. I understand that a POSA would not modify prior art if the potential modification would have made the prior art unfit for the intended purpose. I also understand that an obviousness allegation cannot be supported by a combination of references that would require a substantial reconstruction or redesign of the elements and a change in the fundamental principles that the prior art was designed to operate on.
21. I understand that the obviousness analysis requires considering whether prior art is analogous to the claimed invention such that a POSA would have reasonably considered it and applied its teaching when seeking a solution to the problem identified by the invention.

VI. LEVEL OF ORDINARY SKILL IN THE ART

22. I understand that Petitioner and Dr. Gabriel take the position that a POSA

“at the relevant time frame would have been a person with a (i) Ph.D. in molecular biology, genetics, bioinformatics, or a related field, and at least two years of experience in high-throughput sequencing technologies or (ii) Master’s degree in one of the same fields with at least four years of experience in high-throughput sequencing technologies.” Illumina’s Petition (“Pet.”), 10; EX1005 (Gabriel Decl.) ¶ 36. I disagree that a POSA would have necessarily been on an interdisciplinary team, however. Pet., 10. I believe a sufficient level of education and experience with high-throughput sequencing technologies would have been enough for an individual to have been a POSA in the field of the invention, without a further requirement that they also be on an interdisciplinary team.

23. In either case, I believe I am amply qualified to offer opinions from the perspective of a POSA in this case. I received my Ph.D. from the University of California – Berkley in 1989 in Biophysical Chemistry. As discussed above, and since receiving my Ph.D., I have decades of research and experience in the study of nucleic acids, including the structure function, and sequencing methods of nucleic acids. I worked on an interdisciplinary team as a consultant to Pacific Biosciences, where I helped integrate chemistry, physics, engineering, and biology to create

novel sequencing technologies. Well before the time of the invention, 2009, I was qualified as a person of ordinary skill in the art in the field of the '730 patent. I have considered my analysis from the perspective of a POSA at the time of the invention.

VII. TECHNICAL BACKGROUND

24. A POSA would understand that the '730 patent relates to “methods and compositions for determining genotypes in patient samples” using DNA sequencing technologies. EX1001, 1:23-24; EX1005 (Gabriel Decl.), ¶34 (“The '730 patent generally relates to a method of nucleic acid sequencing which uses tags or barcodes to correct for errors or bias.”).

A. Sanger-Style Sequencing

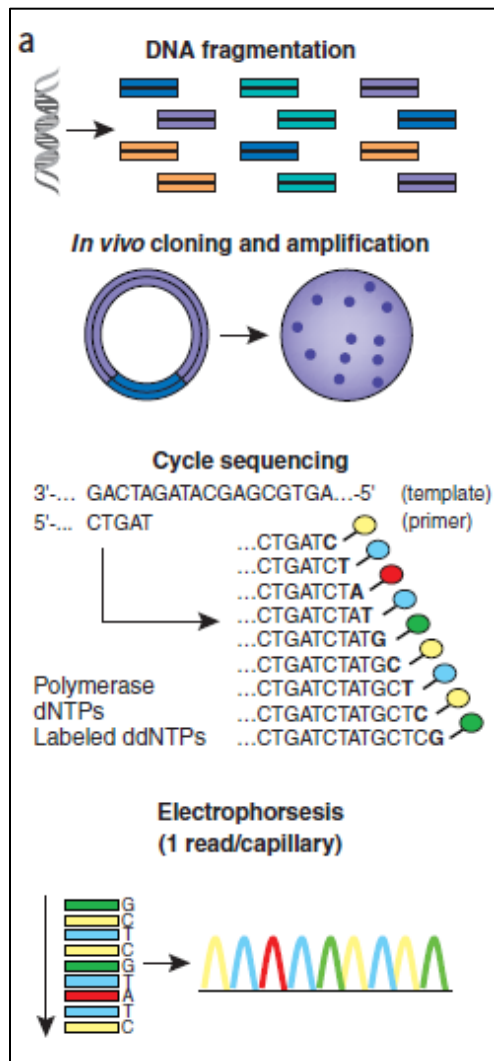
25. In the 1970s, Dr. Fred Sanger developed a DNA sequencing method, commonly known today as Sanger sequencing, that marked a major breakthrough in DNA sequencing technology because it provided one of the first reliable solutions for DNA sequencing. By the 1990s, Sanger sequencing had developed to include “capillary-based, semi-automated implementations” of Sanger sequencing using fluorescence detection in efforts to expand sequencing efficiency and output to support the human genome project. EX1014 (Shendure), 1135.
26. The Sanger sequencing method is known for generating long sequence

reads and is based on the complementary pairing of G with C and A with T. It uses a combination of nucleotide chemistry and an enzyme that copies DNA, called DNA polymerase, which uses one strand of the DNA double helix as a template and “builds” (i.e. polymerizes) the complementary nucleotides using nucleotide triphosphate substrates. EX1014 (Shendure), 1135, 1136. The sequence of the copied strand is complementary to that of the strand sequenced. The Sanger method uses nucleotide analogues that are incorporated into a growing DNA synthesized by DNA polymerase through several consecutive “rounds” or “cycles,” but cannot be further extended beyond the addition of a single nucleotide each round due to the presence of chain terminators. EX1014 (Shendure), 1135. One performs four different reactions with a doped chain terminating in “ddT”, “ddA”, “ddG” and “ddC” nucleotides, and the sequence of the DNA could be read by separating the final DNA products by size using gel electrophoresis. For example, if the template had a G at a given position, the copied DNA will extend and be terminated by a ddC. EX1014 (Shendure), 1135, 1136. This is an early example of sequencing DNA by synthesizing a copy with DNA polymerase (sequencing by synthesis or SBS).

27. A POSA would have understood that each sequencing cycle provides the

identity of a single nucleotide base. EX1014 (Shendure), 1135, 1136. Thus, by obtaining the identity of each nucleotide at each position within a sequence through multiple sequencing cycles, Sanger sequencing provides the sequence of an entire DNA fragment. EX1014 (Shendure), 1135–36.

28. Figure 1 of Shendure, shown below, (EX1014) provides an illustration of Sanger sequencing in the mid-2000s.



EX1014 (Shendure), 1136 (Figure 1).

A POSA would have understood that Sanger sequencing requires multiple cycles of sequencing to provide the read of one DNA fragment per capillary.

29. Sanger sequencing was modernized by applying capillary gel separations and fluorescent chain terminators (allowing sensitive and rapid detection). These technologies were incorporated into the first commercial DNA sequencers, marketed by Applied Biosystems among others in the 1980s. The technology was further enhanced by the efforts of the Human Genome Project in the 1990s, providing much of the information in the landmark human genome sequences that took more than a decade to acquire, with the project completing in 2003. The Sanger method provided (and still does provide) highly accurate sequencing (around 99.99%) of 200-300 nucleotide fragments of DNA (the read length of the technology). It is still the most accurate form of sequencing, and remains widely used.
30. However, the Sanger method suffers from drawbacks including limited throughput (bases sequenced per run or unit time), time consuming sample preparation, and the cost per nucleotide sequenced. These aspects were the major drivers for innovation in sequencing technology in the

early 2000s.

31. For example, by 2008, Sanger sequencing machines had developed “[s]imultaneous electrophoresis in 96 or 364 independent capillaries” that provided “a limited level of parallelization.” EX1014 (Shendure), 1135. But the cost and time associated with Sanger sequencing provided a barrier to many individual researchers. EX1014 (Shendure), 1143 (“The reduction in costs of DNA sequencing by several orders of magnitude is democratizing the extent to which individual investigators can pursue projects at a scale previously accessible only to major genome centers.”), 1141 (“[L]arge-scale resequencing studies for identifying germline variation or somatic mutations have relied on Sanger-based resequencing approaches, that in turn are reliant on *one-at-a-time* PCR amplification of each targeted region. In this context, the requirements of a Sanger sequencing approach *include major costs beyond the reagents*. ... In a recent informal survey we conducted of the overall cost to conventionally sequence 100 genes from 100 samples, assuming each gene has an average of 10 exons quoted estimate from noncommercial genome centers and commercial sequence service providers ranged from \$300,000 to over \$1,000,000. Clearly, *this cost is beyond the range of most individual laboratories*.”).

32. Sanger sequencing methods have since become known as “first generation” sequencing approaches. EX1005 (Gabriel Decl.) ¶ 55.

B. Next Gen Sequencing (NGS)

33. Next-generation sequencing technologies utilize a sequencing-by-synthesis method to sequence millions of short DNA fragments in a single run. EX1017 (Bentley) at 546-548 (“Read lengths [of the Solexa system] are 30-50 bases, which are of sufficient length for re-sequencing applications.”); EX1005 (Gabriel Decl.) ¶ 55, 59 (“[T]he early versions of Next-Gen techniques could generate sequence reads that are 50 to 150 nucleotides (nts) or base pairs (bp) in length.”).
34. By the time of the invention, researchers had developed second generation sequencing methods, or “next-generation” sequencing approaches that “reduc[ed] the cost of DNA sequencing by over two orders of magnitude, and “democratiz[ed] the field by putting the sequencing capacity of a major genome center in the hands of individual investigators.” EX1014 (Shendure), 1135. These sequencing approaches, including the 454 Roche, Solexa or Genome Analyzer, and SOLiD systems, focused on sequencing massive numbers of DNA fragments in parallel. EX1014 (Shendure), 1136-1141; EX1005 (Gabriel Decl.) ¶¶ 55-57.

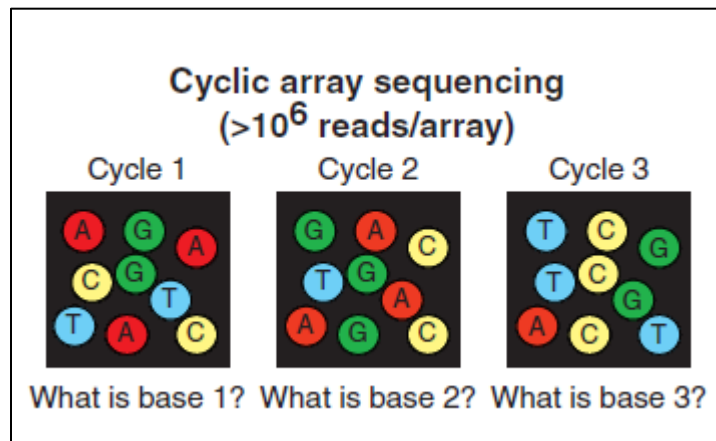
35. Next-generation sequencing approaches “enable a much higher-degree of parallelism than conventional capillary-based sequencing.” EX1014 (Shendure), 1136. For example, “hundreds of millions of sequencing reads can potentially be obtained” with next-generation sequencing when compared to the hundreds of sequencing reads that can be obtained from Sanger sequencing techniques. EX1014 (Shendure), 1136; EX1005 (Gabriel Decl.) ¶ 55.
36. A POSA would thus have understood that next-generation sequencing dramatically reduced the time and cost of DNA sequencing, making it more accessible to individual investigators. EX1014 (Shendure), 1135 (“Over the past three years, massively parallel DNA sequencing platforms have become widely available, reducing the cost of DNA sequencing by over two orders of magnitude, and democratizing the field by putting the sequencing capacity of a major genome center in the hands of individual investigators....Next-generation DNA sequencing has the potential to dramatically accelerate biological and biomedical research, by enabling the comprehensive analysis of genomes, transcriptomes and interactomes to become inexpensive, routine and widespread, rather than requiring significant production-scale efforts.”), 1141 (“In a recent informal survey we conducted of the overall cost to conventionally

sequence 100 genes from 100 samples, assuming each gene has an average of 10 exons quoted estimate from noncommercial genome centers and commercial sequence service providers ranged from \$300,000 to over \$1,000,000. Clearly, this cost is beyond the range of most individual laboratories. In addition to reducing the per-base cost of sequencing by several orders of magnitude, second-generation instruments have fewer infrastructure requirements; instead, the principle challenge is downstream data management.”), 1144 (“The reduction in the costs of DNA sequencing by several orders of magnitude is democratizing the extent to which individual investigators can pursue projects at a scale previously accessibly only to major genome centers.”).

37. Next-generation sequencing generally involves a library preparation process, a sequencing process, and a sequencing analysis process. EX1014 (Shendure), 1136; EX1005 (Gabriel Decl.) ¶¶ 58-74. The library preparation process involves preparing nucleic acids for sequencing. EX1005 (Gabriel Decl.) ¶ 58. For DNA sequencing, the process usually involves obtaining short fragments of DNA and attaching adaptor sequences to prepare the fragments for subsequent sequencing. EX1005 (Gabriel Decl.) ¶¶ 58-60; EX1014 (Shendure), 1136. A POSA would have understood that some next-generation sequencing approaches do

not require PCR amplification.

38. The sequencing process involves performing millions of sequence reads in parallel within a single sequencing run. EX1005 (Gabriel Decl.) ¶ 55. Many next-generation systems use cyclic array sequencing, where multiple DNA fragments are attached to a position on an array and then read in parallel through cyclic addition of nucleotides in a single sequence. EX1014 (Shendure), 1136; EX1005 (Gabriel Decl.).



EX1014 (Shendure), 1136 (Figure 1). Thus, a POSA would have understood that next-generation sequencing allowed for millions of sequence reads to be obtained in a single sequencing run. At the time, a POSA would have understood that 2nd generation sequencing reads were often less than 100 nts.

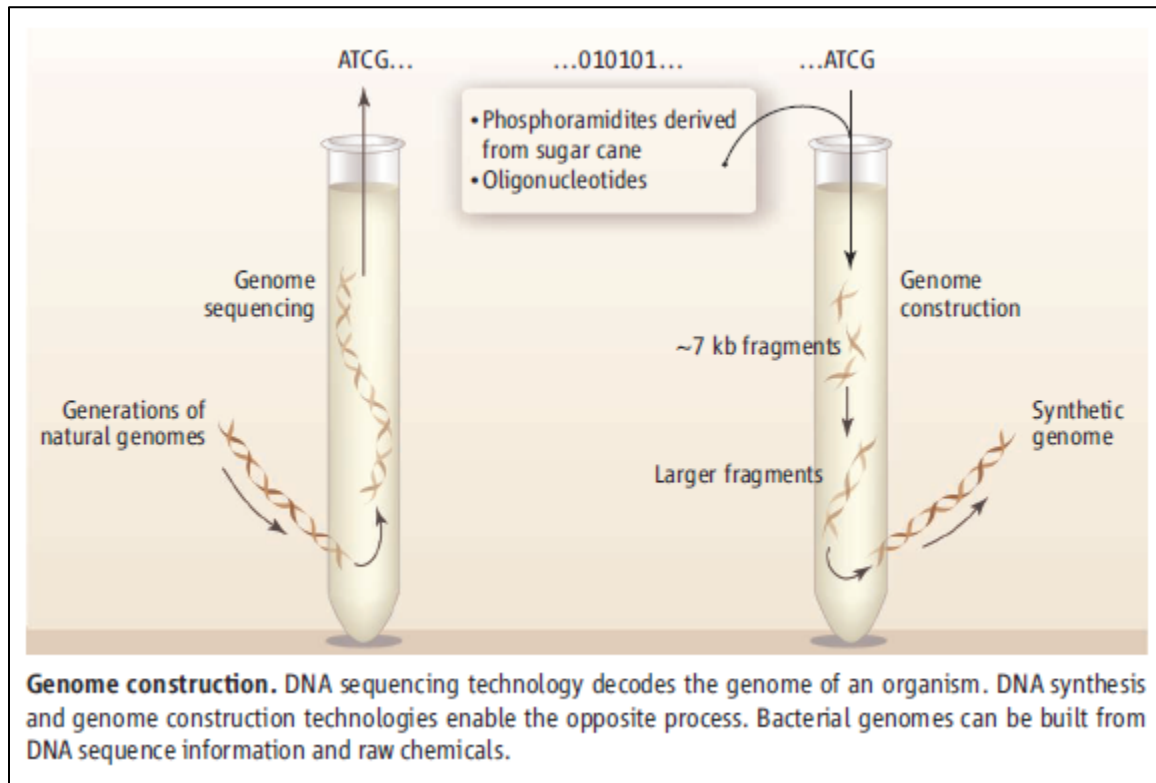
39. The millions of short sequence reads are then used for analysis. For example, computer software and algorithms can align the reads to reference genomes and reconstruct DNA molecules. See EX1005

(Gabriel Decl.) ¶¶ 70-71; EX1014 (Shendure), 1141-1143. A POSA would have understood that there was “significant innovation in developing new algorithms specifically tailored for short reads” because many of the algorithms designed for long reads “are unlikely to be the best algorithms for handling short-read sequence data.” EX1014 (Shendure), 1142. Thus, a POSA would have understood that “the principal challenge [for next-generation sequencing] is downstream data management.” EX1014 (Shendure), 1141.

C. Synthetic Biology

40. Synthetic biology relates to the synthetic creation of biological materials, including nucleic acid sequences such as DNA sequences, to manipulate biology. EX1010 (Chee), 1:11-31 (“Synthetic biology is a promising new field at the interface of engineering and biology. Recently parallel methods have been used to synthesize many different DNA sequences simultaneously, and to provide them as a pool of oligonucleotides.”) (citing EX2003 (Endy)); EX2019 (Keasling), 65 (“Synthetic biology is the design and construction of new biological components, such as enzymes, genetic circuits, and cells, or the redesign of existing biological systems.”). As depicted in the figure below, unlike “DNA sequencing technology” designed to “decode[] the genome of an organism,”

synthetic biology relates to “DNA synthesis and genome construction technologies [which] enable the opposite process.” EX2003 (Endy), 1196.



41. The field of synthetic biology, which grew up in the early 2000s, proposes to engineer biological systems beyond the simpler genetic engineering described above. Synthetic biology attempts to use principles of both synthetic chemistry and those of engineering to harness the power of organisms to provide insights and perform practical tasks. In one aspect of synthetic biology, genes or collections of genes are reengineered. EX2003 (Endy), 1196. For example, Keasling and co-

workers engineered a biosynthetic pathway (multiple genes) to improve the synthesis of the important antimalarial drug artemisinin. EX2018 (Keasling 2006), 940. Another spectacular application of synthetic biology was the achievement of the Venter Institute in the full synthesis for synthetic DNAs of an entire synthetic genome. EX2020 (Hutchison), 3. In other aspects, the genetic control and regulatory aspects of an organism are engineered using principles from electric circuits. Organisms can be turned into sensing machines, detecting pathogens or small molecules using engineered nucleic acid systems, such as aptamers. Finally, synthetic biology can create artificial DNAs and RNAs, with novel rules of complementarity, expanding the repertoire of what can be achieved chemically by an organism.

42. At the time of the invention, many of these approaches were still being developed and applied. In 2008, researchers in the synthetic biology field had recently synthetically constructed the genome of the *Mycoplasma genitalium* bacteria, which was the “first construction of a genome encoding a self-reproducing organism.” EX2003 (Endy), 1196. However, the breakthrough construction “highlight[ed] the need for improved DNA construction technology” and “require[d] highly skilled experts and considerable resources” to perform. EX2003 (Endy), 1196,

1997. For example, the “construction of multi-kilobase scale DNA molecules remain[ed] difficult and costly, largely because [] de novo chemical synthesis of the DNA building blocks is quite error-prone,” making the “intrinsic error rate of solid-phase oligo synthesis ... a major limitation in the cost effective assembly of larger DNA molecules.” EX1010 (Chee), 1:38-43. A POSA would have understood that errors inherent in the process of synthesizing DNA molecules presented a major limitation in synthetic biology. EX1010 (Chee), 1:38-43.

43. The fundamental building block for these synthetic biology experiments was DNA that could reprogram an organism. These experiments often required changing multiple genes, or in the extreme, an entire genome, so long, correct pieces of designed DNA were required. Around the time the inventions were developed, many methods were developed in this time period to create such pieces either from synthetic DNA fragments made in parallel or through polymerase chain reaction amplification often coupled to use of synthetic nucleotides, and various molecule biology tricks to assemble DNA pieces correctly. Chemical synthesis, PCR and molecular manipulations can all introduce errors in the final DNA.

D. Genotyping

44. At the time of the invention, next-generation sequencing technologies were becoming increasingly utilized in diagnostic applications. One example of next-generation sequencing was “genomic resequencing” “for somatic or germline variation discovery in individual(s) of a species for which a reference genome is available.” EX1014 (Shendure), 1143. For example, at the time of the invention researchers were using next-generation sequencing technologies to sequence human DNA for diagnostic and risk assessment purposes. *See* EX1001, 3:43-58.
45. The genotype of a sample is the unique inherited genome of that given sample and determining the genotype of a given sample is critical in many diagnostic applications. For example, “a diagnosis, prognosis, or disease risk assessment [can be] provided to a subject based on a genotype determined for that subject at one or more genetic loci.” EX1001, 3:43-47. A POSA would therefore understand that accurate determination of a genotype is critical to the accuracy of many diagnostic next-generation sequencing applications.

VIII. The '730 Patent

A. Background

46. The '730 patent is titled “Methods and Compositions for Evaluating

Genetic Markers.” See EX1001 (’730 Patent), (54). The inventors of the ’730 patent are Gregory Porreca and Uri Laserson. EX1001, (72). The ’730 patent was filed on November 19, 2020, and claims priority to provisional application No. 61/174,470, filed on April 30, 2009. However, as I have noted below in Section VIII.D.E, *infra*, the inventors of the ’730 patent conceived of the invention of the ’730 patent at least as early as March 6, 2009.

47. The ’730 patent “relates to methods and compositions for ***determining genotypes in patient samples.***” EX1001 (’730 Patent), 1:23-24; *see also* EX1001, 1:28-41 (“Information about the genotype of a subject is becoming more important and relevant for a range of healthcare decisions as the genetic basis for many disease, disorders, and physiological characteristics.”), 5:12-33, 29:21-27, 30:48-51. For example, the ’730 patent provides “multiplex diagnostic methods” to analyze nucleic acid sequences from human samples. EX1001 (’730 Patent), 3:1-20.
48. The inventors of the ’730 patent recognized a need for an efficient, cost-effective way to confidently analyze nucleic acid sequences from patient samples. For example, at the time of the ’730 invention in 2009, “[t]he frequency of incorrect or ambiguous calls [was] too high for [the] current technology to manage large numbers of patient samples without

involving expensive and time-consuming steps to resolve uncertainties and provide confidence in the information output.”. EX1001 (’730 Patent), 1:46-53. Additionally, the inventors of the ’730 patent recognized that “certain types of genetic information can be under-represented or over-represented in a genetic analysis due to a combination of stochastic variation and systematic bias” in the genetic analysis workflow. EX1001 (’730 Patent), 2:1-9. The inventors of the ’730 patent further recognized that over-representation and under-representation can result in an incorrect genotype call. EX1001 (’730 Patent), 27:45-28:2.

49. The inventors of the ’730 patent also identified that it is “useful to determine whether target nucleic acid sequences identified in an away are independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule” in order to “obtain sequence information that is reliable and can be used to make a *genotype conclusion* (e.g., call) with a desired degree of confidence.” EX1001 (’730 Patent), 5:11-24.

B. The Invention Of The ’730 Patent Addressed Accuracy Issues With Next-Gen Sequencers

50. The ’730 patent also recognized that the accuracy of next generation

sequencers in 2009 was “relatively low,” and that therefore oversampling the genomic loci of interest was crucial: “Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.” EX1001 (’730 Patent), 32:11-13 (“Considering that the raw accuracy of most next-generation sequencing instruments is relatively low, it is crucial to oversample the genomic loci of interest.”). The ’730 patent notes that such oversampling is important “to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous.” EX1001 (’730 Patent), 32:11-18; *see also* EX1001, 64:52-65:21.

51. At the time of the ’730 patent invention a POSA would have understood that oversampling resulting in an over- and under- representation of sequences was a problem associated with library construction techniques and genome amplification. *See* EX1001 (’730 Patent), 27:45-28:2. Such disproportionate representation of sequences can result in a heterozygous allele being incorrectly evaluated as a homozygous allele, and vice versa. EX1001 (’730 Patent), 27:47-57. As such, the ’730 patent identifies a need for a new approach to accurately genotype samples. EX1001 (’730

Patent), 27:45-28:24.

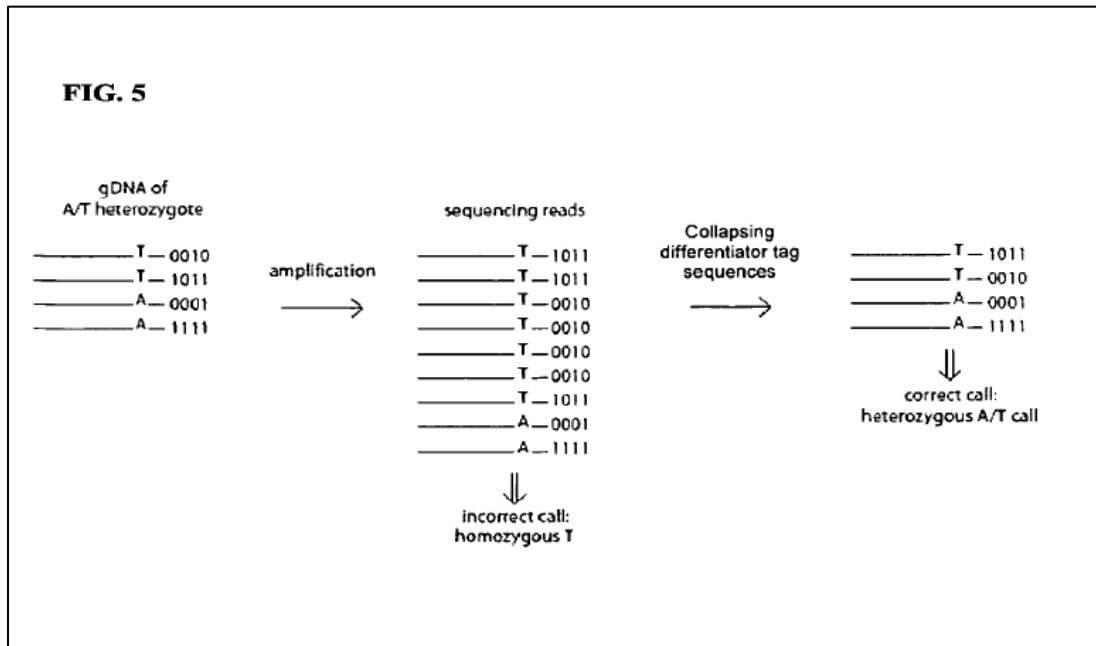
52. The '730 patent discloses the use of “differentiator tag sequences ... to significantly improve the accuracy and confidence of the genotype calls.” EX1001 ('730 Patent), 32:35-38. For example, the '730 patent explains differentiator tags are “useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules.” EX1001 ('730 Patent), 28:29-32. A POSA would have understood that the '730 patent discloses differentiator tags to help overcome disproportionate representation of sequences associated with oversampling. EX1001 ('730 Patent), 5:11-24, 28:25-48, 32:14-38, 64:52-65:21.
53. To address these issues, the claims of the '730 patent provide methods that use differentiator tags to correct for errors or bias introduced during nucleic acid workflows and thereby increase the accuracy and confidence of genotyped samples, particularly in next-gen sequencers. EX1001 ('730 Patent), Claims 1 and 9. For example, the '730 patent explains that differentiator tags “are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g. due to bias in a preparative step, for example,

amplification).” EX1001 (’730 Patent), 5:12-18. The ’730 patent discloses using differentiator tags “to make a genotype conclusion (e.g. call) with a desired degree of confidence” and “detect bias in one or more nucleic acid preparative steps.” EX1001 (’730 Patent), 5:18-24. I summarize the methods described in the ’730 patent below.

54. *First*, nucleic acid molecules (e.g., DNA) that are associated with more than one locus of origin are obtained from any biological sample (e.g., blood). EX1001 (’730 Patent), Claims 1 and 9, 1:42-53, 5:11-24, 13:54-14:3. For example, the ’730 patent explains that “two or more (e.g. 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) different loci may be interrogated in parallel” in a sequencing reaction. EX1001 (’730 Patent), 9:27-29.
55. *Second*, differentiator tags are introduced to and associated with the target nucleic acid molecules. EX1001 (’730 Patent), Claims 1 and 9, 29:7-27, 31:10-32:10. The ’730 patent discloses using “sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule.” EX1001 (’730 Patent), 5:13-23.
56. *Third*, both the tagged target nucleic acid molecules are amplified and sequenced. EX1001 (’730 Patent), Claims 1 and 9.

57. *Finally*, the tagged nucleic acid molecules are analyzed to correct for error and bias introduced during the workflow. EX1001 ('730 Patent), Claims 1 and 9. For example, figure 5 of the '730 patent provides “a non-limiting method for genotyping based on target and differentiator tag sequences” where “reads of target and differentiator tags sequences are collapsed” to correct for errors and bias and make the correct diploid call. EX1001 ('730 Patent), 30:48-51, 72:67-73:5 (“For any target:differentiator tag combination observed more than once, all such reads are ‘collapsed’ into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. FIG. 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination[s] are collapsed.”

Figure 5 “depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination[s] are collapsed.”), (Figure 5 below).



C. Prosecution History of the '730 Patent

58. Dr. Gabriel notes that during the prosecution of the '730 patent, the Examiner rejected the claims as being obvious over three prior art references (Porreca, Wang, and Schiffman). EX1005 (Gabriel Decl.) ¶ 31. In response to the rejection, one of the inventors, Dr. Porreca, submitted a declaration on March 4, 2021. EX1004, 964-967.
59. In his declaration, Dr. Porreca explained that the '730 patent discloses “tagging individual target nucleic acid molecule with differentiator tags before any amplification or sequencing steps” so that “each individual target nucleic acid molecule is tagged.” EX1004, 964. Dr. Porreca further explained that “each target:differentiator tag combination uniquely

identifies a single nucleic acid molecule through subsequent amplification and sequencing steps.” EX1004, 964. Dr. Porreca further explains that the ’730 patent discloses that “[t]arget:differentiator tag combinations observed more than once can then be collapsed into a single read to cancel the effect of bias on consensus basecall accuracy.” EX1004, 965.

60. The Examiner rejected the application and the ’730 patent was allowed after appeal. *See* EX1004 1070-1073, 1142-1146, 1194-1196. In part of its decision allowing the patent, the Board explained that the ’730 patent requires “collapsing, or recognizing, *target:differentiator tag combinations* observed more than once into a single count, or having the same sequence as being derived from the same input molecule, respectively.” EX1004, 1191 (emphasis in original). Because, Porreca (EX1031) does not disclose a “combination of target and differentiator tag” it could not “collapse or recognize target:differentiator tag combination observed more than once.” EX1004, 1191. The claims of the ’730 patent were allowed after the appeal on September 28, 2023. EX1004, 1194-1196.

D. The '730 Patent Is Entitled To A Priority Date Of At Least April 29, 2009

61. I understand that a reference is only prior art under pre-AIA 35 U.S.C. § 102(a) if it was “described in a printed publication ... before the invention thereof by the applicant for a patent.” I further understand under pre-AIA 35 U.S.C. § 102 that the invention of a patent can be established by proof of conception and reasonable diligence in reducing to practice.
62. I understand that conception is the formation, in the mind of the inventor, of a definite and permanent and permanent idea of the complete and operative invention. I understand that conception occurs when the inventor’s idea is definite and permanent enough that a POSA could understand the invention. I understand that the inventor or inventors do not need to know that their invention will work for the conception to be complete. I understand that proof of conception may come from a single source or multiple sources. I understand that proof of conception must include every element of the claimed invention and can be met with corroborated inventor testimony.
63. While I understand that reasonable diligence is required, I was not asked to address diligence.

64. Based on my review of the materials, for the reasons explained in the following Section VIII.E, the inventors conceived of the claimed invention in claims 1-3 and 9 by at least March 6, 2009. EX2006 (March 6 Draft Application); EX2007; EX2008.

E. A POSA Would Have Understood That The March 6, 2009 Documents Discloses Claims 1-3 and 9 Elements Of The '730 Patent

65. On March 6, 2009, Gregory Porreca, one of the inventors on the '730 patent, emailed the start of a draft of a patent application (the "March 6 Application") along with figures for the patent application that eventually became the '730 patent, to his prosecuting attorneys. EX2006 (March 6 Application); EX2007; EX2008. The metadata of the draft application confirms that the document was created by at least March 5, 2009. EX2009. The metadata of the figures confirms that the figures were created by at least March 6, 2009. EX2010; EX2011.
66. A POSA would have understood that the disclosures of the March 6, 2009 application demonstrate conception of claims 1-3, and 9 of the '730 patent, as I discuss further below.

1. Claim 1

67. The preamble of claim 1 recites "[a] method for correcting errors or bias introduced during nucleic acid analysis workflow, the method

comprising the steps of.” EX1001, 131:28-30. The March 6 Application was titled “Methods for Reducing Bias in Preparations of Nucleic Acids” and “relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 (March 6 Draft Application), at 1. The March 6 Application also describes multiple methods to reduce bias in nucleic acid analysis workflows. EX2006 (March 6 Draft Application), 2-5. A POSA would have understood that the March 6 Application discloses “[a] method for correcting for correcting errors or bias introduced during nucleic acid analysis workflow” as stated in the ’730 patent. EX1001, 131:28-30.

68. Claim 1 of the ’730 patent requires “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 131:31-33. The March 6 Application describes methods for DNA and RNA sequencing workflows involving genotyping, genome resequencing, and gene expression, among others. EX2006 (March 6 Application), at 1, 4-5. For example, the March 6 Application further describes examples using “genome re-sequencing” where multiple target nucleic acid molecules are analyzed using next-generation sequencing. EX2006 (March 6 Application), at 4-5 (“One

example of the use of the barcodes is in genome resequencing. ... [S]ince there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to confidently determine whether the sample is homozygous or heterozygous.”). A POSA would have understood the March 6 Application is intended for next-generation sequencing and one advantage of next-generation sequencing is the ability to sequence multiple loci of origin efficiently and inexpensively. EX2006 (March 6 Application), 1, 4-5. Therefore, a POSA have understood that the March 6 Application discloses “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 131:31-33.

69. Claim 1 of the '730 patent also requires “introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag.” EX1001, 131:34-38. The March 6 Application provides methods that use “a large library of unique molecular barcodes in preparative steps for reducing the effects of [] bias.” EX2006 (March 6 Application), at 2. The March 6 Application describes various methods to incorporate barcodes

with target nucleic acids. EX2006 (March 6 Application), at 3-4. The March 6 Application describes methods where “multiple reads have been collected for a given locus” and “a large fraction of the reads derive from a single originating molecule.” EX2006 (March 6 Application), at 4-5. The March 6 Application also explains that “there are two alleles at every locus” and that “it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to confidently determine whether the sample is homozygous or heterozygous.” EX2006 (March 6 Application), at 4. The figures accompanying the March 6 Application also disclose “introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag.” EX2006 (March 6 Figures). For example, the figures show multiple genes with multiple differentiator tags attached to each gene. Thus, a POSA would have understood that the March 6 Application and its figures discloses “introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag.” EX1001, 131:34-38.

70. Claim 1 of the '730 patent requires “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.” EX1001, 131:39-40. The March 6 Application describes methods to overcome biases such as bias caused by amplification. EX2006 (March 6 Application), at 4 (“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads derive from a single originating molecule. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. ... However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule).”). The figures accompanying the March 6 Application disclose amplifying the differentiator tag and the target nucleic acid to generate amplicons. EX2006 (March 6 Figures). Thus, a POSA would have understood that the March 6 Application and accompanying figures disclose “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.”
71. Claim 1 of the '730 patent requires “sequencing the amplicons obtained

in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 131:42-46. The March 6 Application discloses methods using next-generation sequencing to sequence target nucleic acid sequences and differentiator tags for various applications such as genotyping, gene expression, and genome re-sequencing. EX2006 (March 6 Application), at 1, 4-5. The figures accompanying the March 6 Application also disclose sequencing amplicons containing target nucleic acid sequences and a differentiator tag. EX20007; EX2008. Thus, a POSA would have understood that the March 6 Application and its figures disclose “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 131:42-46.

72. Claim 1 of the '730 patent also requires “correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.” EX1001, 131:47-49. The March 6 Application explains that biases and errors can be introduced “into the sequencing libraries, generally due to biases in

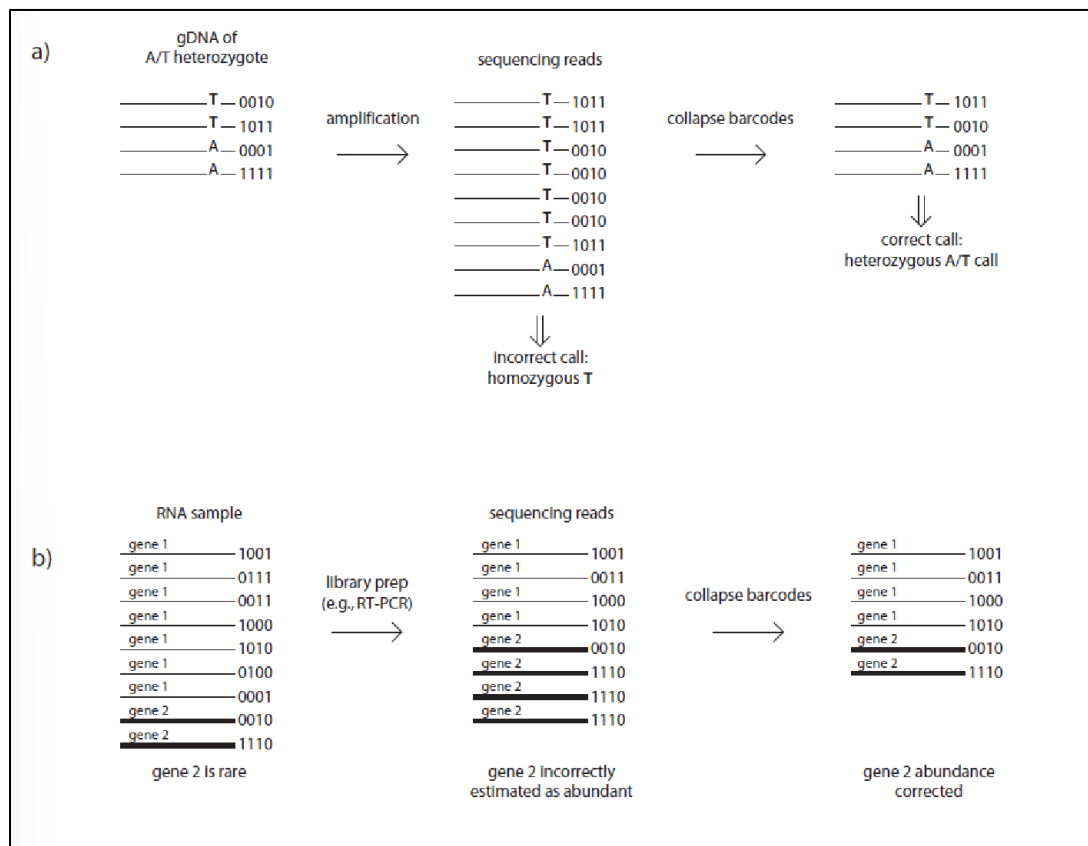
enzyme function or to the exponential and stochastic nature of amplification methods like PCR.” EX2006 (March 6 Application), at 1. The March 6 Application further identifies issues with over- and under-representation of alleles in genotyping applications. EX2006 (March 6 Application). The figures accompanying the March 6 Application also disclose collapsing the target nucleic acid and the differentiator tag into a single count to correct for errors or bias. EX2007; EX2008. For example, the figures show target nucleic acid and differentiator tag combinations getting collapsed into a single count. Thus, a POSA would have understood that the March 6 Application and its figures disclose “correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.” EX1001, 131:47-49.

73. A POSA would have understood that the March 6 Application discloses all of the limitations in claim 1 in detail sufficient for a POSA to understand the claimed invention.

2. Claim 2

74. Claim 2 requires the elements of claim 1 and further requires “collapsing all sequence reads before determining a consensus base call for the target nucleic acid molecule sequences.” EX1001, 131:47-49. A POSA would

have understood that the March 6 Application and the accompanying figures discloses collapsing all sequence reads before determining a consensus base call for the target nucleic acid molecule sequences. For example, the figures accompanying the March 6 Application depict sequence reads collapsed to determine a consensus base call.



EX2008.

75. A POSA would have understood that the March 6 Application discloses all of the limitations in claim 2 in detail sufficient for a POSA to understand the claimed invention.

3. Claim 3

76. Claim 3 requires the elements of claim 2 and further requires “determining the presence of a sequencing or amplification error based upon a number of said differentiator tags.” EX1001, 131:54-57. The March 6 Application explains that “observations of the random barcode sequences are used to *detect and correct* biases in the final read-out of the library.” EX2006 (March 6 Application) at 2. A POSA would have understood that the use of unique barcodes could help determine the presence of sequencing or amplification error. For example, the March 6 Application identifies bias due to amplification processes and provides methods to detect and correct those biases. EX2006 (March 6 Application) at 4-5. The figures accompanying the March 6 Application show how differentiator tags can be used to detect and correct for amplification and sequencing errors. EX2007; EX2008. Thus, a POSA would have understood that the March 6 Application discloses “determining the presence of a sequencing or amplification error based upon a number of said differentiator tags.” EX1001, 131:54-57.
77. A POSA would have understood that the March 6 Application discloses all of the limitations in claim 3 in detail sufficient for a POSA to understand the claimed invention.

4. Claim 9

78. The preamble of Claim 9 requires “[a] method for correcting for errors or bias introduced during nucleic acid analysis workflow.” EX1001, 132:38-40. Because the preamble of claim 9 does not differ from the preamble of claim 1, for the reasons stated above a POSA would have understood that the March 6 Application and its accompanying figures disclose the preamble of claim 9. Section VIII.E.1.
79. Claim 9 requires “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 132:41-43. Because the “obtaining” element of claim 9 does not differ from the “obtaining” element of claim 9, for the reasons stated above a POSA would have understood that the March 6 Application and its accompanying figures disclose the “obtaining” step of claim 9. Section VIII.E.1.
80. Claim 9 requires “introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus or origin.” EX1001, 132:44-48. The March 6 Application provides methods that use “a large library of unique molecular barcodes in preparative steps for reducing the

effects of [] bias.” EX2006 (March 6 Application) at 2. The March 6 Application describes various methods to incorporate barcodes with target nucleic acids. EX2006 (March 6 Application) at 3-4. The March 6 Application discloses using a random and unique barcodes that “are randomly incorporated” with individual molecules. EX2006 (March 6 Application) at 2-3. Thus, a POSA would have understood that the March 6 Application and its figures discloses “introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus or origin.” EX1001, 132:44-48.

81. Claim 9 requires “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.” EX1001, 132:49-50. Because the “amplifying” element of claim 9 does not differ from the “amplifying” element of claim 9, for the reasons stated above a POSA would have understood that the March 6 Application and its accompanying figures disclose the “amplifying” step of claim 9. Section VIII.E.1.
82. Claim 9 requires “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of

the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 132:51-55. Because the “sequencing” element of claim 9 does not differ from the “sequencing” element of claim 9, for the reasons stated above a POSA would have understood that the March 6 Application and its accompanying figures disclose the “sequencing” step of claim 9. Section VIII.E.1.

83. Claim 9 requires “correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.” EX1001, 132:56-59. The March 6 Application explains that biases and errors can be introduced “into the sequencing libraries, generally due to biases in enzyme function or to the exponential and stochastic nature of amplification methods like PCR.” EX2006 (March 6 Application) at 1. The March 6 Application further identifies issues with over- and under-representation of alleles in genotyping applications. EX2006 (March 6 Application). The March 6 Application explains that “[t]he incorporation of the barcode at a given step allows the detection and correction of biases in any of the” sequencing workflow steps. EX2006 (March 6 Application) at 2. The March 6 Application further explains that “any barcodes or barcode-molecule combinations” can detect and correct for

“bias in the protocol between the incorporation of the random barcodes and the actual sequencing.” EX2006 (March 6 Application) at 3. The figures accompanying the March 6 Application further depict “correcting for errors or bias” by “recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.” EX2007; EX20008; EX1001, 132:56-59. Thus a POSA would have understood that the March 6 Application and its accompanying figures disclose “correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.” EX1001, 132:56-59.

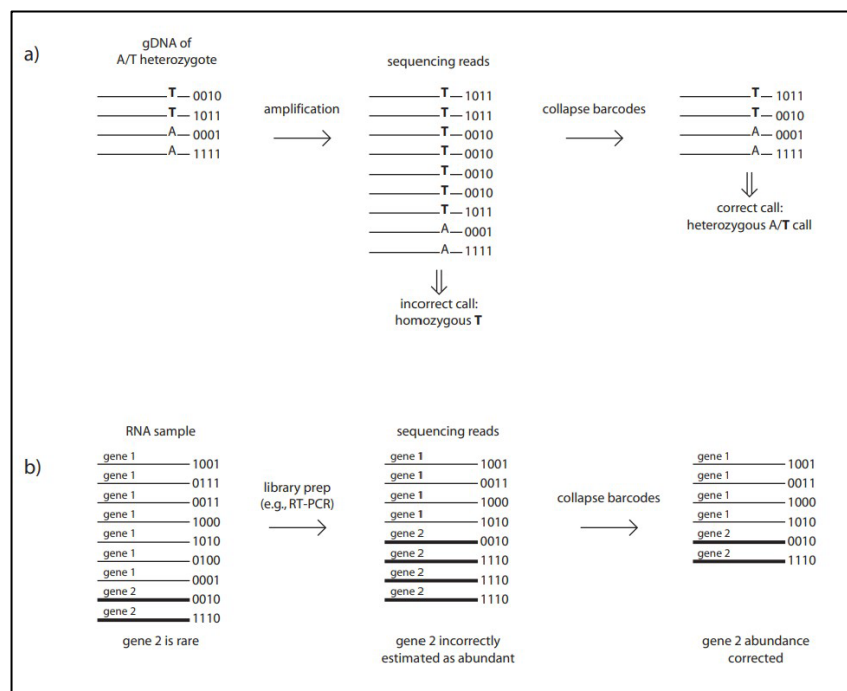
84. A POSA would have understood that the March 6 Application discloses all of the limitations in claim 9 in detail sufficient for a POSA to understand the claimed invention.

Claim of the '730 Patent	Proof of Conception By March 6, 2009
Claim 1, preamble: “A method for correcting errors or bias introduced during nucleic acid analysis workflow , the method comprising the steps of:”	“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1. “ A common problem to overcome is that these techniques commonly introduce biases into the sequencing libraries, generally due to biases in enzyme function or to the exponential and stochastic nature of amplification methods like PCR.” EX2006, at 1. “The incorporation of the barcode at a given step allows the detection and correction of biases in any of the following steps of the protocol.” EX2006, at 2.

“In the final readout of the *sequencing process*, each read has an additional unique barcode sequence. ... If any barcodes or barcode-molecule combinations are statistically overrepresented, it is a sign that there was *some kind of bias in the protocol between the incorporation of the random barcodes and the actual sequencing.*” EX2006, at 3.

“Using the random barcode method, the distribution of the random barcodes is well known at the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after the sequencing is performed, *we have collected the information that is necessary to correct the biases.* By undoing the distribution changes that are measured for the random barcodes, the corresponding changes in the counts for the sequences of interest will provide a *higher fidelity representation* of the composition of the original complex pool.” EX2006, at 3.

“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to *significantly improve the accuracy and confidence of the genotype calls.*” EX2006, at 5.



<p>Claim 1(a): “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;”</p>	<p>EX2008.</p> <p>“There are currently many genomics assays that <i>utilize next-generation (polony-based) sequencing to generate data</i>, including <i>genome resequencing, RNA-seq for gene expression</i>, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling and gene expression profiling), these methods utilize the <i>counts of sequencing reads of a given genomic locus</i> as a proxy for the representation of that <i>sequence in the original sample of nucleic acids</i>” EX2006 at 1.</p> <p>“One example of the use of the barcodes is in genome resequencing. ... Furthermore, since there are <i>two alleles</i> at <i>every locus</i>, it is important to sample enough to ensure that <i>both alleles</i> have been observed a sufficient number of times to confidently determine whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after <i>multiple reads</i> have been <i>collected for a given locus</i>, it is possible that due to bias (e.g., caused by PCR amplification steps), <i>a large fraction of the reads derive from a single originating molecule</i>. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 4-5.</p>
--	---

	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>amplification</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>library prep (e.g., RT-PCR)</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p> <p>gene 2 is rare</p> <p>EX2008.</p>
<p>Claim 1(b):</p> <p>“introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;”</p>	<p>“The invention is the composition and use of a large library of <i>unique molecular barcodes</i> in preparative steps for reducing the effects of the bias described above. During the library preparation process, <i>each molecule is tagged</i> with a <i>unique [barcode]</i> in such a configuration such that the barcode is sequenced along with the information of interest for which the library is being prepared.” EX2006 at 2.</p> <p>“One possible method to create a large library of <i>unique barcodes</i> is to use degenerate, random-sequence barcodes of defined length. The sequences of the barcodes will be read at the final stage of the sequencing. The observations of the random barcode sequences are used to detect and correct biases in the final read-out of the library. The total possible number of such random barcodes is 4^N, where N is the length of the barcode. The length of the barcode is adjusted such that the size of the population of unique barcodes causes any given molecule-barcode combination to be very rare; since <i>the barcodes are randomly incorporated</i>, the probability that two individual molecules receive the same barcode sequence is very small. (In fact, the likelihood that the same molecule barcode-combination occurs more than once is even smaller.)” EX2006 at 2.</p> <p>“In the final readout of the sequencing process, each read has an additional <i>unique barcodes</i> sequencing.” EX2006 at 2.</p>

	<p>“This allows the <i>random barcodes</i> to be used in <i>any preparative process</i> where the ultimate output is sequencing of <i>many molecules in parallel</i>.” EX2006 at 4.</p> <p>“One example of the use of the barcodes is in <i>genome-resequencing</i>.” EX2006 at 4.</p> <p>“[A]fter <i>multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads derive from a single originating molecule</i>. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. . . . However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule).” EX2006 at 4.</p> <div data-bbox="625 934 1404 1554"> <p>a)</p> <p>b)</p> </div> <p>EX2008.</p>
<p>Claim 1(c): “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;”</p>	<p>“The invention relates to preparative steps in <i>DNA sequencing-related technologies</i> that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1.</p> <p>“All of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of the sample of interest. This may include chemical or biochemical</p>

	<p>treatment of the DNA (e.g., reverse transcription, bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a <i>variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification)</i>.” EX2006 at 1.</p> <p>“A common problem to overcome is that these techniques commonly introduce biases into the sequencing libraries, generally due to biases in enzyme function or to <i>exponential and stochastic nature of amplification methods like PCR</i>.” EX2006 at 1.</p> <p>“The number of reads of a given sequence is typically taken as a proxy for the amount of that sequence present in the originating sample. . . . However, as the preparative steps taken during the protocol are known to cause biases and changes in the distribution (particularly for relatively rare events), the fidelity of using the count numbers is often questionable. Using the random barcode method, the distribution of the random barcodes is well known at the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after the sequencing is performed, we have collected the information that is necessary to correct the biases.” EX2006 at 3.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., <i>caused by PCR amplification steps</i>), a <i>large fraction of the reads derive from a single originating molecule</i>. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 4-5.</p>
--	---

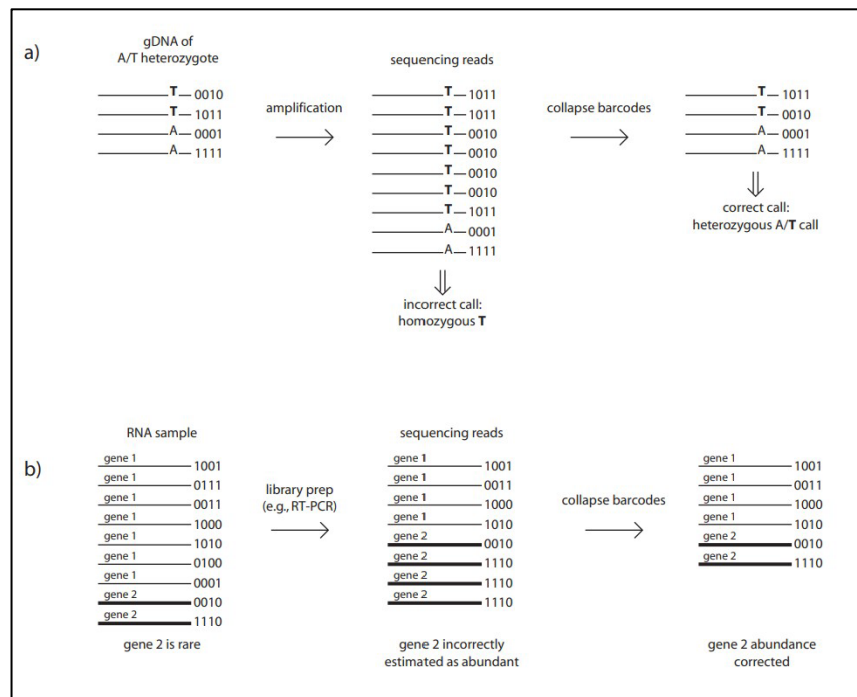
	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>library prep (e.g., RT-PCR)</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p>
<p>Claim 1(d):</p> <p>“sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and”</p>	<p>EX2008.</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1.</p> <p>“There are currently many genomics assays that utilize <i>next-generation (polony-based) sequencing</i> to generate data, including <i>genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq</i>, among others. In order to make quantitative measurements (including <i>genotype calling and gene expression profiling</i>), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX2006 at 1.</p> <p>“During the library preparation process, each molecule is tagged with a unique [barcode] in such a configuration such that <i>the barcode is sequenced along with the information of interest for which the library is being prepared.</i>” EX2006 at 2.</p> <p>“In the final readout of the <i>sequencing</i> process, <i>each read has an additional unique barcode sequence.</i> Since the barcodes were distributed randomly, we should expect that all the unique barcode sequences should be observed an equal number of times. ... If barcodes or <i>barcode-molecule combinations</i> are statistically overrepresented, it is a sign that there was some kind of bias in the</p>

protocol between the incorporation of the random barcodes and the actual sequencing.” EX2006 at 2-3.

“The sequencing primers must be incorporated afterwards such that they *sequence both the barcode and the molecule of interest.*” EX2006 at 3.

“This allows the random barcodes to be used in any preparative process where the ultimate output is *sequencing* of many molecules in parallel.” EX2006 at 3.

“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., *the sequences with the overrepresented barcode all originated from the same single molecule*). Therefore, *the sequences and corresponding distribution of random barcodes* can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 5.



EX2008.

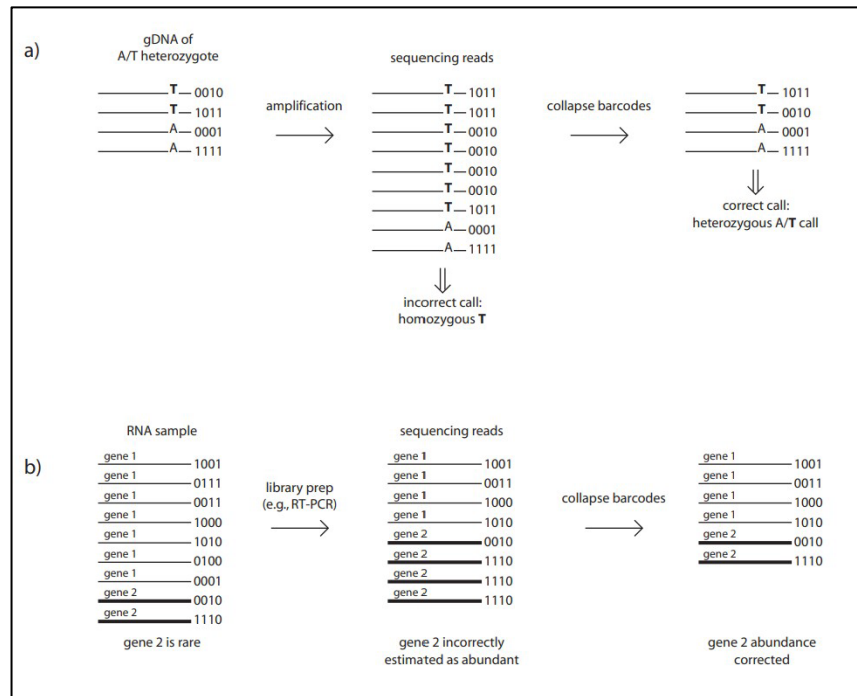
Claim 1(e):
“correcting for error or bias introduced during

“The invention relates to preparative steps in DNA sequencing-related technologies that *reduce bias* and *increase the reliability* and accuracy of downstream quantitative applications.” EX2006 at 1.

<p>said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.”</p>	<p>“A common problem to overcome is that these techniques commonly <i>introduce biases into the sequencing libraries</i>, generally due to <i>biases in enzyme function or to the exponential and stochastic nature of amplification methods like PCR.</i>” EX2006 at 1.</p> <p>“The invention is the composition and use of a large library of unique molecular barcodes in preparative steps <i>for reducing the effects of the bias</i> described above.” EX2006 at 2.</p> <p>“During the library preparation process, each molecule is tagged with a unique [barcode] in such a configuration such that <i>the barcode is sequenced along with the information of interest</i> for which the library is being prepared.” EX2006 at 2.</p> <p>“The incorporation of the barcode at a given step allows the <i>detection and correction of biases</i> in any of the following steps of the protocol.” EX2006 at 2.</p> <p>“[B]y <i>observing how the distribution of random barcodes changes after the sequencing is performed</i>, we have collected the information that is necessary to <i>correct the biases</i>. By undoing the distribution changes that are measured for the random barcodes, the corresponding changes in the counts for the sequences of interest will provide <i>a higher fidelity representation of the composition of the original complex pool.</i>” EX2006 at 3.</p> <p>“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, <i>the sequences and corresponding distribution of random barcodes</i> can be used as an additional input to the genotype-calling algorithm to <i>significantly improve the accuracy and confidence of the genotype calls.</i>” EX2006 at 5.</p>
---	---

	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p>
<p>Claim 2:</p> <p>“The method of claim 1, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.”</p>	<p>EX2008.</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that <i>reduce bias</i> and <i>increase the reliability and accuracy of downstream quantitative applications</i>.” EX2006 at 1.</p> <p>“There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including <i>genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq</i>, among others. In order to make <i>quantitative measurements (including genotype calling and gene expression profiling)</i>, these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX2006 at 1.</p> <p>“One example of the use of the barcodes is in <i>genome resequencing</i>.” EX2006 at 4.</p> <p>“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an <i>additional input to the genotype-calling algorithm to significantly</i></p>

improve the accuracy and confidence of the genotype calls.”
EX2006 at 5.



EX2008.

Claim 3:
“The method of claim 2, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.”

“During the library preparation process, each molecule is tagged with a unique [barcode] in such a configuration such that the barcode is sequenced along with the information of interest for which the library is being prepared. The *incorporation of the barcode at a given step allows the detection and correction of biases in any of the following steps of the protocol.*” EX2006 at 2.

“One possible method to create a large library of unique barcodes is to use degenerate, random-sequence barcodes of defined length. The sequences of the barcodes will be read at the final stage of the sequencing. The *observations of the random barcode sequences are used to detect and correct biases in the final read-out of the library.*” EX2006 at 2.

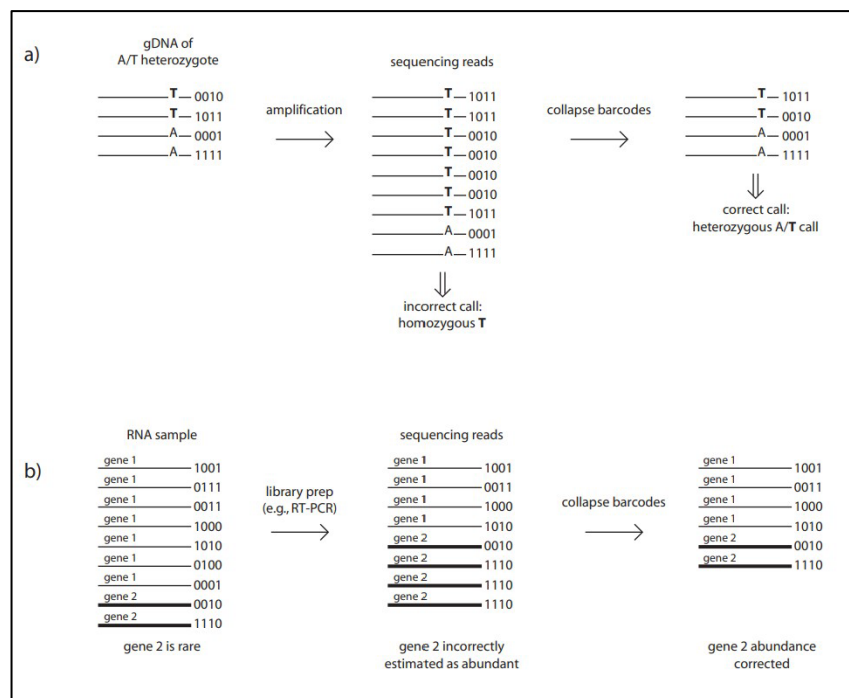
“If any barcodes or barcode-molecule combinations are statistically overrepresented, *it is a sign that there was some kind of bias in the protocol between the incorporation of the random barcodes and the actual sequencing.*” EX2006 at 3.

“Using the random barcode method, the distribution of the random barcodes is well known at the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after

the sequencing is performed, *we have collected the information that is necessary to correct the biases.*” EX2006 at 3.

“At that point, *any biases* in any part of the protocol would change the uniformity of this distribution, which *can be observed after sequencing.*” EX2006 at 4.

“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads derive from a single originating molecule. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 4-5.

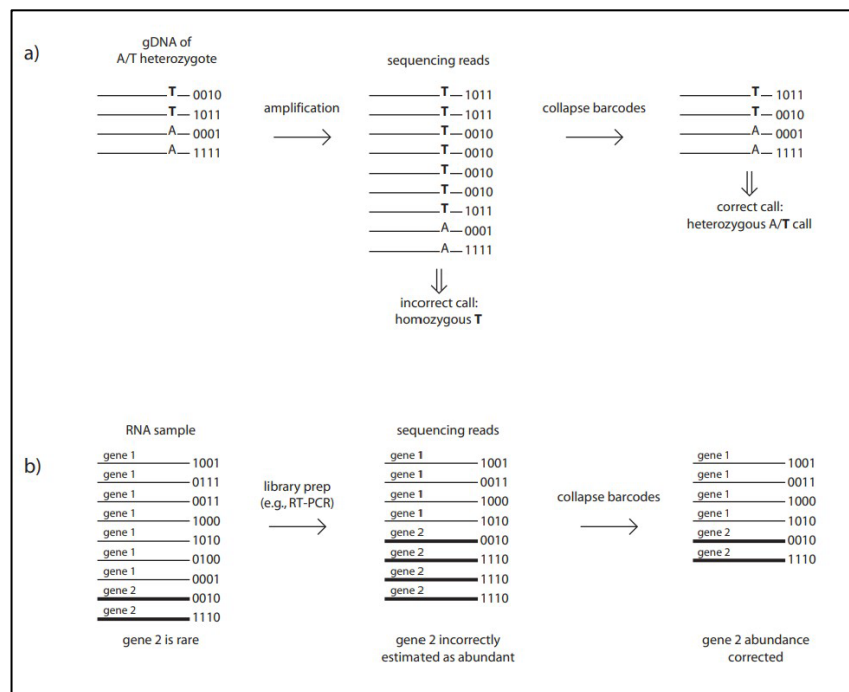


EX2008

<p>Claim 9, Preamble: “A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:”</p>	<p>“The invention relates to <i>preparative steps in DNA sequencing-related technologies</i> that <i>reduce bias</i> and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1.</p> <p>“<i>A common problem to overcome is that these techniques commonly introduce biases</i> into the sequencing libraries, generally due to biases in enzyme function or to the exponential and stochastic nature of amplification methods like PCR.” EX2006 at 1.</p> <p>“The incorporation of the barcode at a given step <i>allows the detection and correction of biases</i> in any of the following steps of the protocol.” EX2006 at 2.</p> <p>“In the final readout of the <i>sequencing process</i>, each read has an additional unique barcode sequence. ... If any barcodes or barcode-molecule combinations are statistically overrepresented, it is a sign that there was <i>some kind of bias in the protocol between the incorporation of the random barcodes and the actual sequencing</i>.” EX2006 at 3.</p> <p>“Using the random barcode method, the distribution of the random barcodes is well known at the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after the sequencing is performed, <i>we have collected the information that is necessary to correct the biases</i>. By undoing the distribution changes that are measured for the random barcodes, the corresponding changes in the counts for the sequences of interest will provide a <i>higher fidelity representation</i> of the composition of the original complex pool.” EX2006 at 3.</p> <p>“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to <i>significantly improve the accuracy and confidence of the genotype calls</i>.” EX2006 at 5.</p>
--	---

	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p>
<p>Claim 9(a): “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;”</p>	<p>EX2008.</p> <p>“There are currently many genomics assays that <i>utilize next-generation (polony-based) sequencing to generate data</i>, including <i>genome resequencing, RNA-seq for gene expression</i>, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling and gene expression profiling), these methods utilize the <i>counts of sequencing reads of a given genomic locus</i> as a proxy for the representation of that <i>sequence in the original sample of nucleic acids</i>” EX2006 at 1.</p> <p>“However, as the preparative steps taken during the protocol are known to cause biases and changes in the distribution (particularly for relatively rare events), the fidelity of using the count numbers is often questionable. Using the random barcode method, the distribution of the random barcodes is well known at the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after the sequencing is performed, we have collected the information that is necessary to correct the biases.” EX2006 at 3.</p> <p>“One example of the use of the barcodes is in genome resequencing. ... Furthermore, since there are <i>two alleles</i> at <i>every locus</i>, it is important to sample enough to ensure that <i>both alleles</i> have been observed a sufficient number of times to confidently determine whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in</p>

the originating sample. However, after *multiple reads* have been *collected for a given locus*, it is possible that due to bias (e.g., caused by PCR amplification steps), *a large fraction of the reads derive from a single originating molecule*. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 4-5.



EX2008.

Claim 9(b):
“introducing a set of *unique differentiator tags*, wherein members of said set of unique differentiator tags are *randomly associated with*

“The invention is the composition and use of a *large library of unique molecular barcodes* in preparative steps for reducing the effects of the bias described above. During the library preparation process, *each molecule is tagged with a unique [barcode]* in such a configuration such that the barcode is sequenced along with the information of interest for which the library is being prepared.” EX2006 at 2.

<p>members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;"</p>	<p>"One possible method to create a large library of <i>unique barcodes</i> is to use degenerate, <i>random-sequence barcodes of defined length</i>. The sequences of the barcodes will be read at the final stage of the sequencing. The observations of the random barcode sequences are used to detect and correct biases in the final read-out of the library. The total possible number of such random barcodes is 4^N, where N is the length of the barcode. The length of the barcode is adjusted such that the size of the population of unique barcodes causes any given molecule-barcode combination to be very rare; <i>since the barcodes are randomly incorporated</i>, the probability that two individual molecules receive the same barcode sequence is very small. (In fact, the likelihood that the same molecule barcode-combination occurs more than once is even smaller.)" EX2006 at 2.</p> <p>"<i>Since the barcodes were distributed randomly</i>, we should expect that all the <i>unique barcode sequences</i> should be observed about an equal number of times." EX2006 at 2.</p> <p>"This allows the <i>random barcodes</i> to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel." EX2006 at 4.</p> <p>"[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads derive from a single originating molecule. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. . . . However, if <i>we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode</i> (i.e., the sequences with the overrepresented barcode all originated from the same single molecule)." EX2006 at 4.</p>
---	---

	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>amplification</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>library prep (e.g., RT-PCR)</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p> <p>gene 2 is rare</p>
<p>Claim 9(c): “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;”</p>	<p>EX2008.</p> <p>“The invention relates to preparative steps in <i>DNA sequencing-related technologies</i> that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1.</p> <p>“All of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of the sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., reverse transcription, bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a <i>variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification)</i>.” EX2006 at 1.</p> <p>“A common problem to overcome is that these techniques commonly introduce biases into the sequencing libraries, generally due to biases in enzyme function or to <i>exponential and stochastic nature of amplification methods like PCR</i>.” EX2006 at 1.</p> <p>“The number of reads of a given sequence is typically taken as a proxy for the amount of that sequence present in the originating sample. . . . However, as the preparative steps taken during the protocol are known to cause biases and changes in the distribution (particularly for relatively rare events), the fidelity of using the count numbers is often questionable. Using the random barcode method, the distribution of the random barcodes is well known at</p>

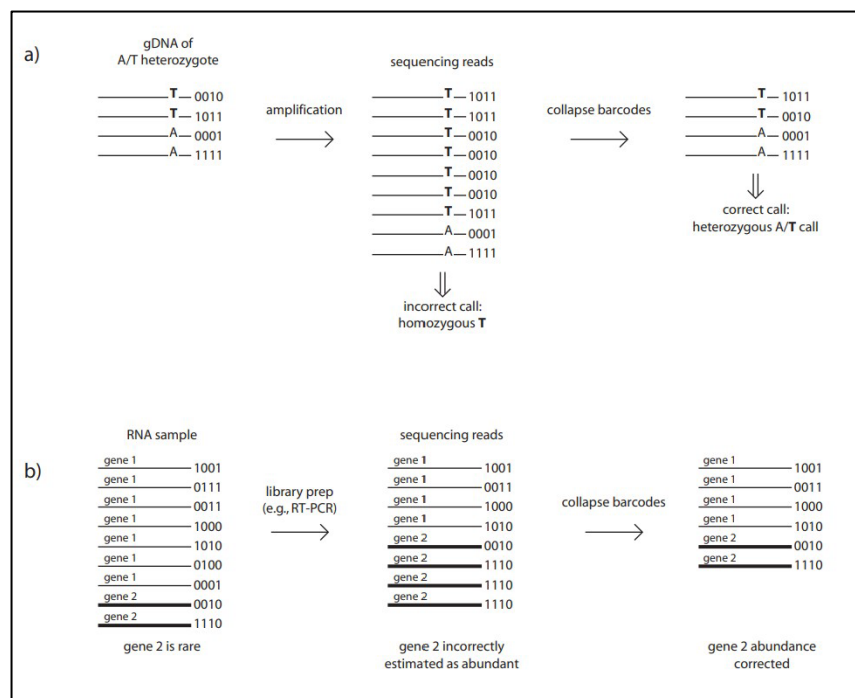
	<p>the time of incorporation (it is uniform); by observing how the distribution of random barcodes changes after the sequencing is performed, we have collected the information that is necessary to correct the biases.” EX2006 at 3.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., <i>caused by PCR amplification steps</i>), a <i>large fraction of the reads derive from a single originating molecule</i>. This would skew the population of genotypes observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 4-5.</p> <div data-bbox="618 1024 1463 1717"> <p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>collapse barcodes</p> <p>gene 2 abundance corrected</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 is rare</p> </div> <p>EX2008.</p>
<p>Claim 9(d): “sequencing the amplicons obtained in said amplifying step to</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX2006 at 1.</p>

obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and”	<p>“There are currently many genomics assays that utilize <i>next-generation (polony-based) sequencing</i> to generate data, including <i>genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq</i>, among others. In order to make quantitative measurements (including <i>genotype calling and gene expression profiling</i>), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX2006 at 1.</p> <p>“During the library preparation process, each molecule is tagged with a unique [barcode] in such a configuration such that <i>the barcode is sequenced along with the information of interest for which the library is being prepared</i>.” EX2006 at 2.</p> <p>“In the final readout of the <i>sequencing</i> process, <i>each read has an additional unique barcode sequence</i>. Since the barcodes were distributed randomly, we should expect that all the unique barcode sequences should be observed an equal number of times. ... If barcodes or <i>barcode-molecule combinations</i> are statistically overrepresented, it is a sign that there was some kind of bias in the protocol between the incorporation of the random barcodes and the actual sequencing.” EX2006 at 2-3.</p> <p>“The sequencing primers must be incorporated afterwards such that they <i>sequence both the barcode and the molecule of interest</i>.” EX2006 at 3.</p> <p>“This allows the random barcodes to be used in any preparative process where the ultimate output is <i>sequencing</i> of many molecules in parallel.” EX2006 at 3.</p> <p>“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., <i>the sequences with the overrepresented barcode all originated from the same single molecule</i>). Therefore, <i>the sequences and corresponding distribution of random barcodes</i> can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX2006 at 5.</p>
--	--

	<p>a) gDNA of A/T heterozygote</p> <p>sequencing reads</p> <p>amplification</p> <p>collapse barcodes</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p> <p>b) RNA sample</p> <p>sequencing reads</p> <p>library prep (e.g., RT-PCR)</p> <p>collapse barcodes</p> <p>gene 2 incorrectly estimated as abundant</p> <p>gene 2 abundance corrected</p>
<p>Claim 9(e):</p> <p>“correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.”</p>	<p>EX2008</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that <i>reduce bias</i> and <i>increase the reliability and accuracy of downstream quantitative applications</i>.” EX2006 at 1.</p> <p>“A common problem to overcome is that these techniques commonly introduce <i>biases into the sequencing libraries</i>, generally due to <i>biases in enzyme function or to the exponential and stochastic nature of amplification methods like PCR</i>.” EX2006 at 1.</p> <p>“The invention is the composition and <i>use of a large library of unique molecular barcodes</i> in preparative steps for <i>reducing the effects of the bias</i> described above.” EX2006 at 2.</p> <p>“During the library preparation process, each molecule is <i>tagged with a unique [barcode] in such a configuration such that the barcode is sequenced along with the information of interest</i> for which the library is being prepared.” EX2006 at 2.</p> <p>“The <i>incorporation of the barcode</i> at a given step allows the <i>detection and correction of biases in any of the following steps of the protocol</i>.” EX2006 at 2.</p> <p>“[B]y observing how the distribution of random barcodes changes after the sequencing is performed, we have <i>collected the</i></p>

information that is necessary to correct the biases. By undoing the distribution changes that are measured for the random barcodes, the corresponding changes in the counts for the sequences of interest will provide a higher fidelity representation of the composition of the original complex pool.” EX2006 at 3.

“However, if we also have information on random barcodes, this situation could be averted, because the over-represented allele would be seen to also have an over-represented random barcode (i.e., the sequences with the overrepresented barcode all originated from the same single molecule). Therefore, *the sequences and corresponding distribution of random barcodes can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.*” EX2006 at 5.



EX2008.

F. A POSA Would have Understood That The May 18, 2009 Provisional Application Discloses Claim 1-3 and 9 Of The '730 Patent.

85. Dr. Gabriel uses the filing date of Provisional Application No.

61/179,358 (the “’358 provisional”), May 18, 2009, as the priority date for the ’730 patent in her analysis. EX1005 (Gabriel Decl.) ¶ 40. Dr. Gabriel asserts that “the earlier provisional applications do not support the claimed subject matter” and uses May 18,2009 as the priority date. EX1005 (Gabriel Decl.) ¶ 40. Based on my review, the ’358 provisional application discloses at least claims 1-3 and 9 of the ’730 patent.

1. Claim 1

86. The preamble of claim 1 recites “[a] method for correcting errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of.” EX1001, 131:28-30. The ’358 provisional application was titled “methods and compositions for reducing diagnostic errors associated with bias in preparations of nucleic acids” and discloses methods to correct for errors and biases in nucleic acid analysis workflows such as “approaches ... for overcoming bias in sequencing library preparatory methods.” EX1008 at 15. For example, the ’358 provisional discloses that its methods “relate[] to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008, at 14. The ’358 provisional also describes multiple methods to reduce bias and errors in nucleic acid analysis

workflows. EX1008 at 9-10, 15, 19-26. A POSA would have understood that the '358 patent discloses the claim “[a] method for correcting errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of.” EX1001, 131:28-30.

87. Claim 1 of the '730 patent requires “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 131:31-33. The '358 provisional application discloses methods for DNA sequencing workflows that involve sequencing human samples at multiple genetic loci. For example, the '358 provisional “disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample” and that “[a] patient or subject may be a human.” EX1008, at 20. Therefore, a POSA would have understood that the '358 provisional discloses “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 131:31-33.

88. Claim 1 of the '730 patent also requires “introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said

loci of origin are associated with more than one differentiator tag.” EX1001, 131:34-38. The ’358 provisional discloses introducing differentiator tags to a plurality of nucleic acids in a nucleic acid analysis workflow. For example, the ’358 provisional describes “associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias.)” EX1008, at 20. The ’358 provisional also explains that “a ‘differentiator tag sequence’ is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality of different isolated nucleic acids, identifies a unique, independently isolated nucleic acid.” EX1008, at 21. Furthermore, the ’358 provisional explains that its methods “allow[] the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel” and “analyz[ing] a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1).” EX1008, at 24, 25. A POSA would have understood that the ’358 provisional discloses “introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are

associated with more than one differentiator tag.” EX1001, 131:34-38.

89. Claim 1 of the '730 patent requires “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.” EX1001, 131:39-40. The '358 provisional explains that its methods “are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important.” EX1008, at 16. The '358 provisional describes methods where “a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample” prior to amplification and that “changes in the occurrence of differentiator tag sequences after amplification and sequence may be indicative of bias in the protocol.” EX1008 at 22-23. A POSA would have understood that the '358 provisional discloses “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.” EX1001, 131:39-40.

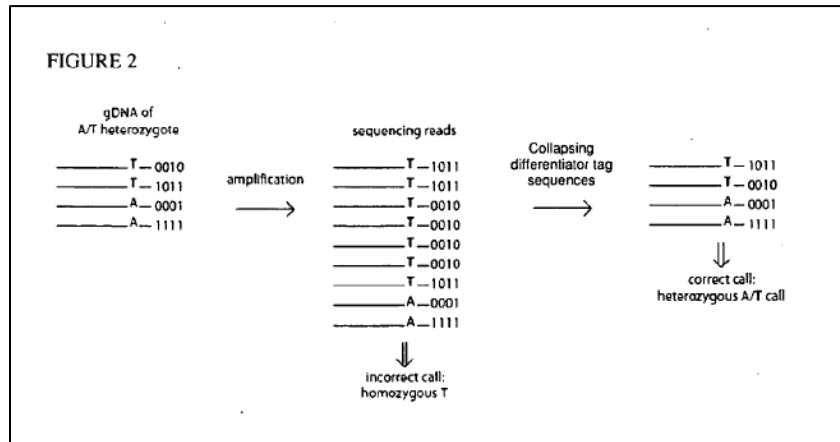
90. Claim 1 of the '730 patent requires “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 131:42-46. The '358 provisional discloses sequencing amplicons containing a

differentiator tag and a target nucleic acid molecule sequence. For example, the '359 provisional explains describes methods where “sequencing reads of target and differentiator tag sequences are colla[se]d to make diploid genotype calls.” EX1008, at 19. The '358 provisional explains that its “invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008, at 14. Furthermore, the '358 provisional explains that “sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence.” EX1008, at 24. Therefore, a POSA would have understood that the '358 provisional discloses “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 131:42-46.

91. Claim 1 of the '730 patent also requires “correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.” EX1001, 131:47-49. The '358 provisional discloses correcting for error or bias in

a nucleic acid analysis workflow by collapsing sequences containing target nucleic acid sequences and differentiator tags into a single count. For example, the '358 provisional explains that its methods "relate[] to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications." EX1008, at 14. The '358 provisional also explains that a "genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process" and that "[s]uch bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries." EX1008, at 14. To address such errors and bias, the '358 provisional explains that "each nucleic acid molecule may be tagged with a unique differentiator tag in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence)." EX1008, at 22. The '358 provisional further discloses that reads with "any target:differentiator tag combination observed more than once ... are 'collapsed' into a single read before consensus basecalls are determined." EX1008, at 44. The

'358 provisional also provides a figure that “depicts a method for genotyping based on target and differentiator tag sequences” where “reads of target and differentiator tag sequences are collapsed to make diploid genotype calls.” EX1008, at 19.

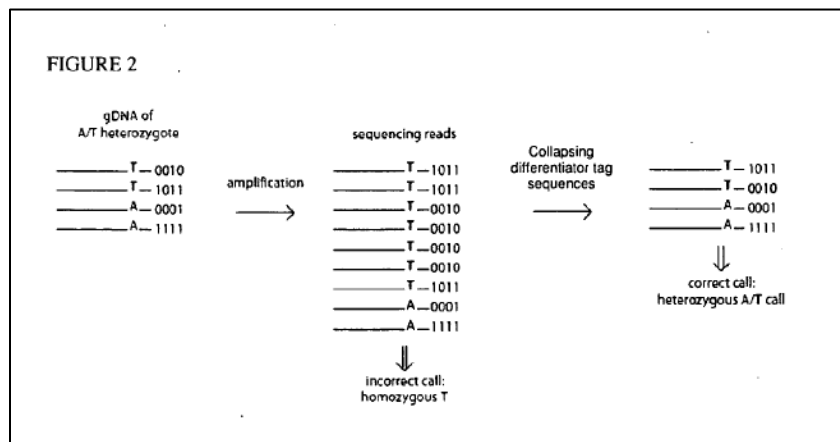


EX1008, at 49 (Figure 2). A POSA would have understood that the '358 provisional discloses “correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.” EX1001, 131:47-49.

2. Claim 2

92. Claim 2 requires the elements of claim 1 and further requires “collapsing all sequence reads before determining a consensus base call for the target nucleic acid molecule sequences.” EX1001, 131:47-49. The '358 provisional discloses collapsing all sequence reads before determining a consensus base call. For example, the '358 provisional explains that reads

with a “target:differentiator tag combination observed more than once” are “‘collapsed’ into a single read before consensus basecalls are determined.” EX1008, at 44. The ’358 provisional explains that the collapsing step “will cancel the effect of bias on consensus basecall accuracy.” EX1008, at 44. The ’358 provisional also provides a figure that “depicts a method for genotyping based on target and differentiator tag sequences” where “reads of target and differentiator tag sequences are collapsed to make diploid genotype calls.” EX1008, at 19.



EX1008, at 49 (Figure 2). A POSA would have understood that the ’358 provisional discloses “collapsing all sequence reads before determining a consensus base call for the target nucleic acid molecule sequences.” EX1001, 131:47-49.

3. Claim 3

93. Claim 3 requires the elements of claim 2 and further requires

“determining the presence of a sequencing or amplification error based upon a number of said differentiator tags.” EX1001, 131:54-57. The ’358 provisional discloses detecting the presence of sequencing or amplification error based upon a number of differentiator tag sequences. For example, the ’358 provisional discloses that “[t]he number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample” and that “changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol.” EX1008 at 22. The ’358 provisional also explains that “[t]he incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol” such as amplification and sequencing. EX1008 at 22. The ’358 provisional further explains that “the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction.” EX1008, at 15. Thus, a POSA would have understood that the ’358 provisional discloses ““determining the presence of a sequencing or amplification error based upon a number of said differentiator tags.” EX1001, 131:54-57.

4. Claim 9

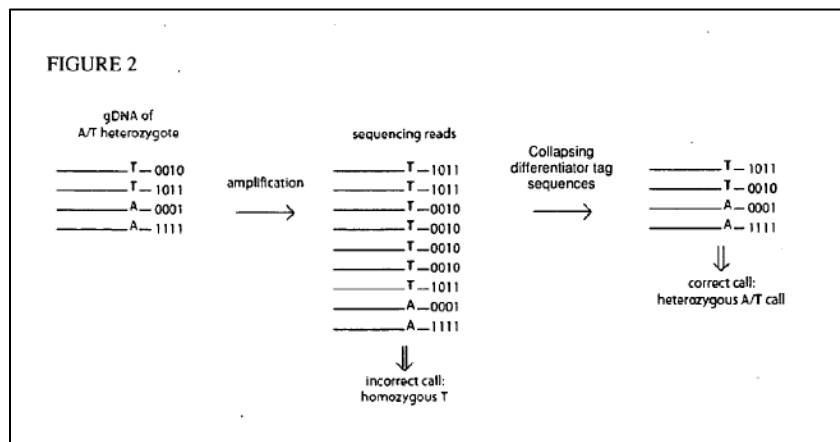
94. The preamble of Claim 9 requires “[a] method for correcting for errors or bias introduced during nucleic acid analysis workflow.” EX1001, 132:38-40. Because the preamble of claim 9 does not differ from the preamble of claim 1, for the reasons stated above a POSA would have understood that the ’358 provisional discloses the preamble of claim 9. Section VIII.F.1.
95. Claim 9 requires “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin.” EX1001, 132:41-43. Because the “obtaining” step of claim 9 does not differ from the “obtaining” step of claim 1, for the reasons stated above a POSA would have understood that the ’358 provisional discloses the “obtaining” step of claim 9. Section VIII.F.1.
96. Claim 9 requires “introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus or origin.” EX1001, 132:44-48. The ’358 provisional discloses introducing unique differentiator tags that are randomly associated with a plurality of target nucleic acid molecules such that any given tag is uniquely associated

with a locus or origin. The '358 provisional explains that “a ‘differentiator tag sequence’ is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality different isolated nucleic acids, identifies a unique, independently isolated nucleic acid” and that “differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method.” EX1008, at 21. The '358 provisional describes using “[a] large library of unique differentiator tag sequences” that can be “distributed randomly in a library.” EX1008 at 22. A POSA would have understood that the '358 provisional discloses “introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus or origin.” EX1001, 132:44-48.

97. Claim 9 requires “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons.” EX1001, 132:49-50. Because the “amplifying” step of claim 9 does not differ from the “amplifying” step of claim 1, for the reasons stated above a POSA would have understood that the '358 provisional discloses the “amplifying” step of claim 9. Section VIII.F.1.

98. Claim 9 requires “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence.” EX1001, 132:51-55. Because the “sequencing” step of claim 9 does not differ from the “sequencing” step of claim 1, for the reasons stated above a POSA would have understood that the ’358 provisional discloses the “sequencing” step of claim 9. Section VIII.F.1.
99. Claim 9 requires “correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.” EX1001, 132:56-59. The ’358 provisional discloses that “[s]ystematic and random errors are common problems associated with genome amplification and sequencing library construction” and that “new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 14-15. The ’358 provisional explains that its methods “relate to associated unique sequence tags (referred to as differentiator tag sequences) with individual molecules that are independently captured and/or analyzed.” EX1008 at 20. The ’358 provisional explains that “tags are useful to distinguish independent target molecules from each other

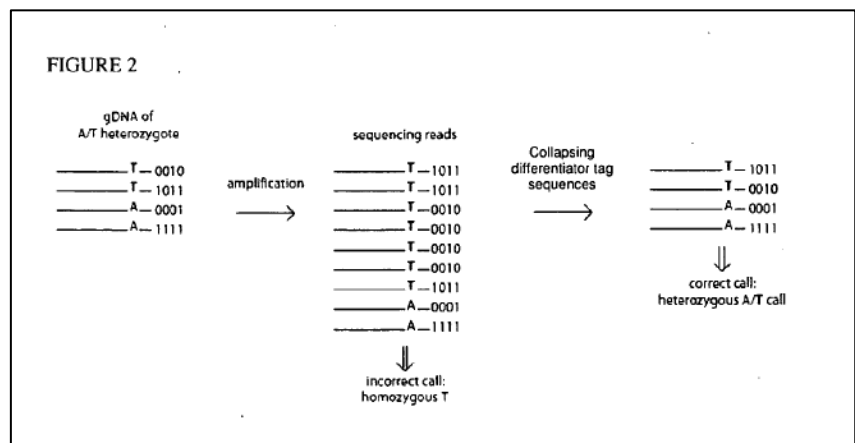
thereby allowing an analysis to be based on a known number of individual target molecules.” EX1008 at 20. The ’358 provisional further provides that “unique differentiator tag sequences ... may be used to detect and correct biases in the final sequencing read-out of the library.” EX1008, at 22. The ’358 provisional also provides a figure where “target:differentiator tag combination[s] observed more than once ... are ‘collapsed’ into a single read before consensus basecalls are determined.” EX1008, at 44.



EX1008 at 49 (Figure 2). A POSA would therefore understand that the ’358 provisional discloses “correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.” EX1001, 132:56-59.

Claim of the '730 Patent	May 18, 2009 Provisional Application
<p>Claim 1, preamble: “A method for correcting errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:”</p>	<p>“The invention relates to <i>preparative steps in DNA sequencing-related technologies</i> that <i>reduce bias</i> and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“<i>Systematic and random errors</i> are common problems associated with <i>amplification and sequencing library construction techniques</i>. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of error (bias) in library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries.” EX1008 at 15.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. ... The number of reads of a given target: differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the number of occurrence of sequences in the original sample is the quantity of interest. ... Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g. for diagnostic purposes.” EX1008 at 22-23.</p>

“However, if we information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to *significantly improve the accuracy and confidence of the genotype calls.*” EX1008, at 12.



EX1008 Figure 2.

Claim 1(a):
“obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;”

“There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest.” EX1008 at 14.

“In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.” EX1008 at 15.

“In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the

	<p>analysis of a biological sample obtained from that subject).” EX1008 at 15.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tag sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“In some embodiments, the disclosure provides methods for analyzing multiple genetic loci (e.g., a plurality of target nucleic acids selected from Table 1) from a patient sample, such as a blood, pre-implantation embryo, chorionic villus or amniotic fluid sample. A patient or subject may be a human. However, aspects of the invention are not limited to humans and may be applied to other species (e.g. mammals, birds, reptiles, other vertebrates or invertebrates) as aspects of the invention are not limited in this respect.”) EX1008 at 20-21.</p> <p>“One example of the use of the differentiator tag sequences is in genome re-sequencing. Furthermore, since there are two alleles at every locus, it is important to sample enough to ensure that both alleles have been observed a sufficient number of times to determine with a sufficient degree of statistical confidence whether the sample is homozygous or heterozygous. Indeed, the sequencing is performed to sample the composition of molecules in the originating sample. However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 24-25.</p>
Claim 1(b):	“In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined

<p>“introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;”</p>	<p>for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject).” EX1008 at 15.</p> <p>“Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.” EX1008 at 20.</p> <p>“As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:diffentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids.” EX1008 at 21.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of</p>
---	---

	<p>target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as ‘target:differentiator tag sequences.’” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence.” EX1008 at 22.</p> <p>“This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.” EX1008 at 24.</p> <p>“One example of the use of the differentiator tag sequences is in genome re-sequencing.” EX1008 at 24.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating</p>
--	---

	<p>molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. ...However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <p>“In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). “ EX1008 at 25.</p> <div data-bbox="639 842 1481 1278"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "Collapsing differentiator tag sequences" --> C["T-1011 T-0010 A-0001 A-1111"] C --> D["correct call: heterozygous A/T call"] B --> E["incorrect call: homozygous T"] </pre> </div> <p>EX1008 Figure 2.</p>
<p>Claim 1(c): “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;”</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification). Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques.” EX1008 at 14.</p>

	<p>“The methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important.” EX1008 at 16.</p> <p>“Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids.” EX1008 at 19.</p> <p>“The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the number of occurrence of sequences in the original sample is the quantity of interest. ...Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g. for diagnostic purposes.” EX1008 at 22-23.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p>
--	---

	<p>FIGURE 2</p> <p>gDNA of A/T heterozygote</p> <p>amplification</p> <p>sequencing reads</p> <p>Collapsing differentiator tag sequences</p> <p>correct call: heterozygous A/T call</p> <p>incorrect call: homozygous T</p>
<p>Claim 1(d): “sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and”</p>	<p>EX1008 Figure 2.</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 14.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence.” EX1008 at 22.</p>

	<p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.</p> <p>“The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence.” EX1008 at 24.</p> <p>“However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <div data-bbox="639 1318 1481 1755"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "↓ incorrect call: homozygous T" --> C["incorrect call: homozygous T"] B -- "Collapsing differentiator tag sequences" --> D["T-1011 T-0010 A-0001 A-1111"] D -- "↓ correct call: heterozygous A/T call" --> E["correct call: heterozygous A/T call"] </pre> </div>
<p>Claim 1(e): “correcting for error or bias introduced during</p>	<p>EX1008 Figure 2.</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p>

<p>said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.”</p>	<p>“Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.” EX1008 at 15.</p> <p>“Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.” EX1008 at 15.</p> <p>“In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. The methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-</p>
---	--

	<p>circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.” EX1008 at 15-16.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information.” EX1008 at 20.</p> <p>“As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:diffentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids.” EX1008 at 21.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p>
--	--

	<p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library.” EX1008 at 22.</p> <p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.</p> <p>“In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.” EX1008 at 24.</p> <p>“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the</p>
--	---

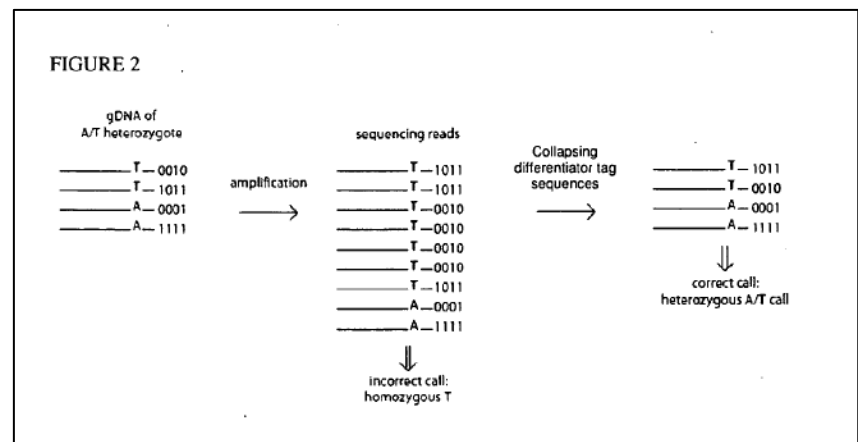
	<p>genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <p>“For any target differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. Figure 2 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.” EX1008 at 44.</p> <div data-bbox="639 989 1481 1423"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "Collapsing differentiator tag sequences" --> C["T-1011 T-0010 A-0001 A-1111"] C --> D["correct call: heterozygous A/T call"] B --> E["incorrect call: homozygous T"] </pre> </div> <p>EX1008 at 49 (Figure 2).</p>
<p>Claim 2: “The method of claim 1, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.”</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX1008 at 14.</p>

	<p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p> <p>“One example of the use of the differentiator tag sequences is in genome re-sequencing.” EX1008 at 24.</p> <p>“However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <p>“In some embodiments, the genotyping comprises determining the sequence of a target nucleic acid (e.g., a polymorphic sequence) at one or more (both) alleles of a genome (a diploid genome) of a subject. In certain embodiments, the genotyping comprises determining the sequence of a target nucleic acid at both alleles of a diploid genome of a subject, wherein in the target nucleic acid comprises, or consists of, a sequence of Table 2.” EX1008 at 18-19.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tag sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“For any target differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. Figure 2 depicts a method for making diploid genotype calls in which repeat target: differentiator tag combination are collapsed.” EX1008 at 44.</p>
--	--

	<p>FIGURE 2</p> <p>The diagram illustrates a sequencing workflow for an A/T heterozygote. It starts with 'gDNA of A/T heterozygote' represented by four horizontal lines: T-0010, T-1011, A-0001, and A-1111. An arrow labeled 'amplification' leads to 'sequencing reads', which are listed as: T-1011, T-1011, T-0010, T-0010, T-0010, T-0010, T-1011, A-0001, and A-1111. From the reads, two paths emerge. One path, indicated by a downward arrow, leads to an 'incorrect call: homozygous T'. The other path, indicated by a rightward arrow labeled 'Collapsing differentiator tag sequences', leads to a final set of four lines: T-1011, T-0010, A-0001, and A-1111. A downward arrow from this final set leads to the 'correct call: heterozygous A/T call'.</p>
<p>Claim 3: “The method of claim 2, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.”</p>	<p>EX1008 at 49 (Figure 2).</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.” EX1008 at 15.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008, at 21-22.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments,</p>

the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.

“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.



EX1008 at 49 (Figure 2).

Claim 9, Preamble:
“A method for correcting
for errors or bias

“The invention relates to *preparative steps in DNA sequencing-related technologies* that *reduce bias* and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.

<p>introduced during nucleic acid analysis workflow, the method comprising the steps of:"</p>	<p>“<i>Systematic and random errors</i> are common problems associated with <i>amplification and sequencing library construction techniques</i>. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of error (bias) in library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries.” EX1008 at 14.</p> <p>“Accordingly,, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. ... The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the number of occurrence of sequences in the original sample is the quantity of interest. ...Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g. for diagnostic purposes.” EX1008 at 22-23.</p> <p>“However, if we information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25 .</p>
---	--

	<p>FIGURE 2</p> <p>The diagram illustrates a method for DNA sequencing that uses differentiator tags to identify individual molecules. It begins with 'gDNA of A/T heterozygote' represented by four horizontal lines with labels: T-0010, T-1011, A-0001, and A-1111. An arrow labeled 'amplification' points to 'sequencing reads', which are shown as nine horizontal lines with labels: T-1011, T-1011, T-0010, T-0010, T-0010, T-0010, T-1011, A-0001, and A-1111. A downward arrow from the reads points to an 'incorrect call: homozygous T'. Another arrow labeled 'Collapsing differentiator tag sequences' points to a final set of four horizontal lines with labels: T-1011, T-0010, A-0001, and A-1111. A downward arrow from this set points to a 'correct call: heterozygous A/T call'.</p>
<p>Claim 9(a): “obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;”</p>	<p>EX1008 Figure 2.</p> <p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.” EX1008 at 15.</p> <p>“Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence</p>

	<p>information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.” EX1008 at 15.</p> <p>“In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. The methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.” EX1008 at 15-16.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information.” EX1008 at 20.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p>
--	--

	<p>“A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library.” EX1008 at 22.</p> <p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.</p> <p>“In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.” EX1008 at 24.</p> <p>“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly</p>
--	---

	<p>improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <p>“For any target differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. Figure 2 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.” EX1008 at 44.</p> <div data-bbox="639 588 1481 1024"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "Collapsing differentiator tag sequences" --> C["T-1011 T-0010 A-0001 A-1111"] C --> D["correct call: heterozygous A/T call"] B --> E["incorrect call: homozygous T"] </pre> </div> <p>EX1008 Figure 2.</p>
<p>Claim 9(b): “introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;”</p>	<p>“In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject).” EX1008 at 15.</p> <p>“In certain embodiments, the disclosure provides preparative methods for capturing target nucleic acids (e.g., genetic loci) that involve the use of differentiator tag sequences to uniquely tag individual nucleic acid molecules.” EX1008 at 19.</p> <p>“Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information. In contrast, if each of the plurality of target molecule sequences</p>

	<p>obtained in the assay is associated with the same differentiator tag, then they probably all originated from the same target molecule due to over-representation (e.g., due to biased amplification) of this target molecule in the assay. This provides less information than the situation where each nucleic acid was associated with a different differentiator tag. In some embodiments, a threshold number of independently isolated molecules (e.g., unique combinations of differentiator tag and target sequences) is analyzed to determine the genotype of a subject.” EX1008 at 20.</p> <p>“As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:diffentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids.” EX1008 at 21.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p>
--	---

	<p>“A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing. The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library. For example, the total possible number of differentiator tag sequences, which may be produced, e.g., randomly, is 4^N, where N is the length of the differentiator tag sequence. Thus, it is to be understood that the length of the differentiator tag sequence may be adjusted such that the size of the population of MIPs having unique differentiator tag sequences is sufficient to produce a library of MIP capture products in which identical independent combinations of target nucleic acid and differentiator tag sequence are rare. As used herein combinations of target nucleic acid and differentiator tag sequences, may also be referred to as ‘target:differentiator tag sequences.’” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence. In some embodiments, when differentiator tag sequences are distributed randomly in a library, all the unique differentiator tag sequences will be observed about an equal number of times.” EX1008 at 22.</p> <p>“This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.” EX1008 at 24.</p> <p>“One example of the use of the differentiator tag sequences is in genome re-sequencing.” EX1008 at 24.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. ...However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p>
--	--

	<p>“In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1).” EX1008 at 25.</p> <div data-bbox="639 405 1481 842"> <p>FIGURE 2</p> <p>The diagram illustrates the process of collapsing differentiator tag sequences. It starts with 'gDNA of A/T heterozygote' showing four sequences: T-0010, T-1011, A-0001, and A-1111. An 'amplification' step leads to 'sequencing reads' which are: T-1011, T-1011, T-0010, T-0010, T-0010, T-1011, A-0001, and A-1111. An arrow points down to an 'incorrect call: homozygous T'. A 'Collapsing differentiator tag sequences' step leads to a list of four sequences: T-1011, T-0010, A-0001, and A-1111. An arrow points down to a 'correct call: heterozygous A/T call'.</p> </div> <p>EX1008 Figure 2.</p>
<p>Claim 9(c): “amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;”</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. This may include chemical or biochemical treatment of the DNA (e.g., bisulphite treatment), capture of a specific subset of the genome (e.g., padlock probe capture, solution hybridization), and a variety of amplification techniques (e.g., polymerase chain reaction, whole genome amplification, rolling circle amplification). Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques.” EX1008 at 14.</p> <p>“The methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important.” EX1008 at 16.</p> <p>“Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and random errors (bias) in genome capture, amplification and sequencing methods, namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries. Accordingly, in some embodiments, the disclosure provides methods that reduce variability in the capture and amplification of nucleic acids.” E EX1008 at 19.</p>

	<p>“The number of reads of a given target:differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the number of occurrence of sequences in the original sample is the quantity of interest. ...Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g. for diagnostic purposes.” EX1008 at 22-23.</p> <p>“[A]fter multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <div data-bbox="639 1281 1481 1717"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "Collapsing differentiator tag sequences" --> C["T-1011 T-0010 A-0001 A-1111"] C --> D["correct call: heterozygous A/T call"] B --> E["incorrect call: homozygous T"] </pre> </div> <p>EX1008 Figure 2.</p>
<p>Claim 9(d): “sequencing the amplicons obtained in said amplifying step to</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p>

<p>obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and”</p>	<p>“There are currently many genomics assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 14.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“In the final readout of a sequencing process, each read may have an additional unique differentiator tag sequence.” EX1008 at 22.</p> <p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.</p>
---	---

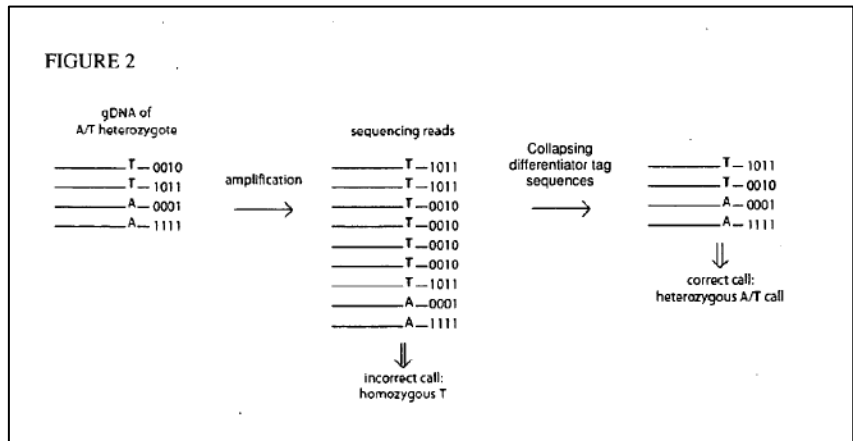
	<p>“The sequencing primers must be incorporated subsequently such that they sequence both the differentiator tag sequence and the target sequence.” EX1008 at 24.</p> <p>“However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p> <div data-bbox="639 770 1481 1205"> <p>FIGURE 2</p> <pre> graph LR A["gDNA of A/T heterozygote T-0010 T-1011 A-0001 A-1111"] -- amplification --> B["sequencing reads T-1011 T-1011 T-0010 T-0010 T-0010 T-0010 T-1011 A-0001 A-1111"] B -- "Collapsing differentiator tag sequences" --> C["T-1011 T-0010 A-0001 A-1111"] C -- "correct call: heterozygous A/T call" --> D["correct call: heterozygous A/T call"] B -- "incorrect call: homozygous T" --> E["incorrect call: homozygous T"] </pre> </div> <p>EX1008, at 49 Figure 2.</p>
<p>Claim 9(e): “correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.”</p>	<p>“The invention relates to preparative steps in DNA sequencing-related technologies that reduce bias and increase the reliability and accuracy of downstream quantitative applications.” EX1008 at 14.</p> <p>“Systematic and random errors are common problems associated with genome amplification and sequencing library construction techniques. For example, genomic sequencing library may contain an over- or under-representation of particular sequences from a source genome as a result of errors (bias) in the library construction process. Such bias can be particularly problematic when it results in target sequences from a genome being absent or undetectable in the sequencing libraries.” EX1008 at 14.</p> <p>“Accordingly, new approaches are needed for overcoming bias in sequencing library preparatory methods.” EX1008 at 15.</p> <p>“In some embodiments, the differentiator tag sequence permit the detection of bias based on the occurrence of combinations of</p>

	<p>differentiator tag and target sequences observed in a sequencing reaction. In other embodiments, the methods reduce errors caused by bias, or the risk of bias, associated with the capture, amplification and sequencing of genetic loci, e.g., for diagnostic purposes.” EX1008 at 15.</p> <p>“Aspects of the invention relate to providing sequence tags (referred to as differentiator tags) that are useful to determine whether target nucleic acid sequences identified in an assay are from independently isolated target nucleic acids or from multiple copies of the same target nucleic acid molecule (e.g., due to bias in a preparative step, for example, amplification). This information can be used to help analyze a threshold number of independently isolated target nucleic acids from a biological sample in order to obtain sequence information that is reliable and can be used to make a genotype conclusion (e.g., call) with a desired degree of confidence. This information also can be used to detect bias in one or more nucleic acid preparative steps.” EX1008 at 15.</p> <p>“In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. The methods disclosed herein are useful for any application where reduction of bias, e.g., associated with genomic isolation, amplification, sequencing, is important. For example, detection of cancer mutations in a heterogeneous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.” EX1008 at 15-16.</p> <p>“Figure 2 depicts a method for genotyping based on target and differentiator tag sequences. Sequencing reads of target and differentiator tags sequences are collapsed to make diploid genotype calls.” EX1008 at 19.</p> <p>“Aspects of the invention relate to associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules that are independently captured and/or analyzed (e.g., prior to amplification or other process that may introduce bias). These tags are useful to distinguish independent target molecules from each other thereby allowing an analysis to be based on a known</p>
--	---

	<p>number of individual target molecules. For example, if each of a plurality of target molecule sequences obtained in an assay is associated with a different differentiator tag, then the target sequences can be considered to be independent of each other and a genotype likelihood can be determined based on this information.” EX1008 at 20.</p> <p>“As used herein, a "differentiator tag sequence" is a sequence of a nucleic acid (a preparative nucleic acid), which in the context of a plurality different isolated nucleic acids, identifies a unique, independently isolated nucleic acid. Typically, differentiator tag sequences are used to identify the origin of a target nucleic acid at one or more stages of a nucleic acid preparative method. For example, in the context of a multiplex nucleic acid capture reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, target nucleic acid capture events. Also, in the context of a multiplex nucleic acid amplification reaction, differentiator tag sequences provide a basis for differentiating between multiple independent, primary amplicons of a target nucleic acid, for example. Thus, combinations of target nucleic acid and differentiator tag sequence (target:diffentiator tag sequences) of an isolated nucleic acid of a preparative method provide a basis for identifying unique, independently isolated target nucleic acids.” EX1008 at 21.</p> <p>“For purposes of genotyping, isolated nucleic acids are identified as independently isolated if they comprise unique combinations of target nucleic acid and differentiator tag sequences, and observance of threshold numbers of unique combinations of target nucleic acid and differentiator tag sequences provide a certain statistical confidence in the genotype.” EX1008 at 21-22.</p> <p>“During a library preparation process, each nucleic acid molecule may be tagged with a unique differentiator tag sequence in a configuration that permits the differentiator tag sequence to be sequenced along with the target nucleic acid sequence of interest (the nucleic acid sequence for which the library is being prepared, e.g., a polymorphic sequence). The incorporation of the nucleic acid comprising a differentiator tag sequence at a particular step allows the detection and correction of biases in subsequent steps of the protocol.” EX1008 at 22.</p> <p>“A large library of unique differentiator tag sequences may be created by using degenerate, random-sequence polynucleotides of defined length. The differentiator tag sequences of the polynucleotides may be read at the final stage of the sequencing.</p>
--	--

	<p>The observations of the differentiator tag sequences may be used to detect and correct biases in the final sequencing read-out of the library.” EX1008 at 22.</p> <p>“The number of reads of a given target differentiator tag sequence may be indicative (may serve as a proxy) of the amount of that target sequence present in the originating sample. In some embodiments, the numbers of occurrence of sequences in the originating sample is the quantity of interest. For example, using the methods disclosed herein, the occurrence of differentiator tag sequences in a pool of MIPs may be predetermined (e.g., may be the same for all differentiator tag sequences). Accordingly, changes in the occurrence of differentiator tag sequences after amplification and sequencing may be indicative of bias in the protocol. Bias may be corrected to provide an accurate representation of the composition of the original MIP pool, e.g., for diagnostic purposes.” EX1008 at 22-23.</p> <p>“In some embodiments, at the incorporation step, the distribution of differentiator tag sequences can be assumed to be uniform. In this case, bias in any part of the protocol would change the uniformity of this distribution, which can be observed after sequencing. This allows the differentiator tag sequence to be used in any preparative process where the ultimate output is sequencing of many molecules in parallel.” EX1008 at 24.</p> <p>“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), a large fraction of the reads are derived from a single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call. For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.” EX1008 at 25.</p>
--	--

“For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy. Figure 2 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination are collapsed.” EX1008 at 44.



EX1008, at 49 (Figure 2).

IX. ILLUMINA’S ASSERTED PRIOR ART

A. Chee

100. Chee (EX1010) is U.S. Patent No. 9,085,798 titled “Nucleic Acid Constructs and Methods of Use.” EX1010 (Chee). Chee relates to synthetic biology, and specifically “a need to select error-free *synthesized oligonucleotides* for use in *synthetic biology* and other analytical methods.” EX1010 (Chee), 1:53-55, *see also* EX1010 (Chee) 1:11-13. Chee was filed on April 29, 2010 as PCT Application No. PCT/US2010/033064. Chee claims priority to provisional application No. 61/174,442, filed on April 30, 2009. EX1010 (Chee), cover.

However, as I have described above, a POSA would have understood that the March 6, 2009 draft application discloses all claim limitations of the '730 patent such that a POSA would have understood the invention of the '730 patent. *See* Section VIII.E. Therefore, I understand that the inventors of the '730 patent conceived of the invention of the '730 patent before Chee's earliest possible priority date.

101. Synthetic biology involves the chemical synthesis of DNA sequences or oligonucleotides. EX1010 (Chee), 1:23-52. Chee explains that “synthesizing relatively small pieces of DNA ... allows a cost-effective production of large amounts of these oligonucleotides. EX1010 (Chee), 1:28-31. Chee explains that use of these small chemically synthesized pieces of DNA allow “much larger segments of DNA” molecules to be constructed such as the entire “*Mycoplasma genitalium* genome.” EX1010 (Chee), 1:28-36.

102. However, Chee further explains that construction of larger “multi-kilobase scale DNA molecules remains difficult and costly, largely because [*the*] *de novo chemical synthesis of the DNA building blocks is quite error prone.*” EX1010 (Chee), 1:38-41 Chee identifies the “intrinsic error rate of solid phase oligo synthesis” as a “major limitation in the cost effective assembly of larger DNA molecules.” EX1010

(Chee), 1:41-43. Therefore, although Chee describes the possibility of performing its methods on DNA or RNA fragments isolated from naturally-occurring samples, Chee's primary purpose is to provide methods to address "a need to select error-free *synthesized* oligonucleotides for use in *synthetic biology*." EX1010 (Chee), 1:53-55.

103. To address that need, Chee's methods specifically provide for the creation and use of "constructs" comprising regions of interest, unique identifiers and "other desired elements." EX1010 (Chee), 12:1-27. These constructs can be synthesized in their entirety, including the region of interest or "ROI," along with the unique identifier and other desired elements. EX1010 (Chee), 12:1-6 ("[T]he construct can be *synthesized as a single unit comprising the ROI*, the unique identifier, and any other desired elements that may be necessary for further analysis, amplification, excision, and/or use of the constructs. The *constructs can be synthesized as a single unit* with all elements provided in a single synthesis step.").

104. Chee explains that it is critical that the ROI is incorporated into the construct in order to achieve the goal of "select[ing] error-free synthesized oligonucleotides." EX1010 (Chee), 2:23-28. For example, Chee explains that "[i]t is a fundamental element of the invention that the

unique identifier is associated with a construct, not the ROI” so that “constructs with ROIs having the *correct sequence* can be differentiated from constructs having *substantially the same ROI but containing errors*.” EX1010 (Chee), 2:23-28.

105. Because the “de novo chemical synthesis” of sequences is “quite error-prone,” Chee explains that “[t]he number of nucleic acid constructs in a set should be large enough to ensure a high probability of obtaining *one correct sequence* or *to confirm a correct sequence* associated with a particular unique identifier.” EX1010 (Chee), 1:38-41, 2:44-47. Therefore, the number of constructs “necessary to ensure the accuracy of the selected ROI or ROIs depends upon” the “error rate of the *synthesis procedures*, the error rate of other methods used *to create constructs, and the like*.” EX1010 (Chee), 13:5-8.

106. In other words, Chee recognizes that the synthesis of DNA molecules is error-prone and its efforts are to obtain and identify at least one perfect copy of that synthetic DNA molecule. EX1010 (Chee), 13:14-19 (“For example, if it were determined that on average one would need 100 copies each of 10,000 ROI sequences in order to have a probability near 1 of obtaining at least 1 perfect copy of each ROI sequence, then the number of constructs that would need to be selected from a master set of

constructs would be targeted to be in the range of 1 million.”). “Chee explains the “use of the uniquely identifiable constructs” ensures that “the determination, selection and/or isolation of error-free regions of interest (ROIs) *are correct.*” EX1010 (Chee), 2:9-14. Chee, therefore, provides methods for “ensuring accuracy of oligonucleotides having a desired sequence.” EX1010 (Chee), 1:55-58.

107. To determine whether a construct with an ROI contains the correct sequence (EX1010 (Chee), 2:23-28), Chee analyzes *multiple clonal copies* of the same construct and makes consensus base calls, which means determining the most common nucleotide that appears at each position of the construct. Consensus base calling results in a consensus sequence (i.e., a consensus base sequence). EX1010 (Chee), 14:42-50. In Chee, *all amplicons* of a given construct are read to determine the most probable base pair in each position in a given ROI, *even if some of those amplicons have identical sequences*. For example, when the same construct containing the exact same ROI and tag is observed more than once, *all clonal copies* of that construct are read to determine the consensus base sequence. EX1010 (Chee), 14:42-50, EX1010 (Chee), 9:34-39 (“Furthermore, because molecules that are identical by descent can be identified, a consensus sequence can be determined using data

from multiple molecules, thereby achieving a much higher accuracy rate than most conventional high throughput methods.”). If there is a difference in the construct’s sequence at a given position in only one copy of that construct, Chee assumes that difference is the result of an error occurring during synthesis or sequencing. EX1010 (Chee), 14:46-50 (“Thus, a variation in the sequence of a single copy of a construct is most likely to be a synthesis or sequencing error if it differs from a consensus sequence derived from multiple clonal copies of the construct, which are identified as identical by descent via the shared unique identifier.”).

108. Thus, in determining the consensus base sequence, Chee follows the most common nucleotide detected at each position of a given construct, which is only decided by *reading multiple clonal copies of constructs*. EX1010 (Chee), 9:34-39, 14:46-50 (“Thus, a variation in the sequence of a single copy of a construct is most likely to be a synthesis or sequencing error if it differs from a consensus sequence derived from multiple clonal copies of the construct, which are identified as identical by descent via the shared unique identifier.”). Determining a consensus base sequence by reading multiple amplicons descending from the same construct aligns with Chee’s focus on using tags to confirm the correct sequence of an ROI with a high degree of accuracy. EX1010 (Chee), 9:40-47 (“These

methods can be especially useful for applications involving large numbers of sequences that need to be determined, such as genomic sequencing. The improvement in accuracy resulting from the ability to determine whether ROIs are identical by descent provides a more sensitive method of detecting a rare sequence variant, such as a low frequency mutation; or a minor allelic variant or haplotype in a sample containing DNA from multiple individuals.”).

109. By reading multiple clonal copies of the same construct Chee can determine whether an ROI contains the desired sequence with a high degree of accuracy. EX1010 (Chee), 13:21-29 (“Following the selection of a set of constructs from a master set, the selected constructs undergo a limited amplification to provide additional material for determining accuracy of the ROI and the associated unique identifier that corresponds to an ROI. In certain aspects, the subset of constructs undergoes at least a 10-fold amplification, and more preferably a 100- to 1000-fold amplification. In some cases, even higher levels of amplification may be needed, e.g., a 10,000 to 100,000-fold amplification.”), 14:39-50 (“Although all existing sequencing mechanisms have an error rate, which varies by platform and according to other key variables, this can be overcome using the methods of the invention. Multiple instances of the

same molecule can be sequenced, and the ROI of an individual parent construct can be identified by its unique identifier, as molecules with identical identifiers arise via amplification of the same original construct. Thus, a variation in the sequence of a single copy of a construct is most likely to be a synthesis or sequencing error if it differs from a consensus sequence derived from multiple clonal copies of the construct, which are identified as identical by descent via the shared unique identifier.”).

B. Gahan

110. Gahan (EX1015) is an article titled “Circulating Nucleic Acids in Plasma and Serum: Recent Developments” by Peter B. Gahan and R. Swaminathan. EX1015 (Gahan). Gahan was published in 2008 and claims to report findings presented during the fifth international conference on circulating nucleic acids in plasma and serum held on August 23-26, 2007, in Moscow, Russia. EX1015 (Gahan), 1.
111. Gahan relates to the “biology of circulating nucleic acids.” EX1015 (Gahan), 2. Gahan describes using different types of DNA isolation kits to isolate cell-free nucleic acids from biological samples. EX1015 (Gahan), 3. Gahan explains that isolating cell-free DNA, such as fetal DNA or tumor DNA, can improve diagnostic methods. EX1015 (Gahan), 3, 4. However, Gahan explains that, “[w]hile some workers have made

an attempt to determine the kit with the greatest yield and the largest range of fragment sizes for both RNA and DNA, there is still a lack of clarity concerning the use of the broad range of kits and their applicability.” EX1015 (Gahan), 3. Gahan also notes that, “the methodology is far from consistent, and if there are to be reliable protocols for diagnostic purposes, there needs to be a form of quality assurance for these procedures.” EX1015 (Gahan), 3.

112. Gahan does not disclose uniquely tagging each individual nucleic acid molecule isolated from a biological sample. Gahan also does not describe any mechanism of amplification or sequencing, using next-generation sequencers or otherwise. Gahan also does not describe any methods related to isolating synthetically created nucleic acids.

C. McCloskey

113. McCloskey (EX1012) is U.S. Patent Application Publication No. 2007/0020640 titled “Molecular Encoding of Nucleic Acid Templates for PCR and Other Forms of Sequence Analysis.” EX1012 (McCloskey). McCloskey was published on January 25, 2007.
114. McCloskey was considered by the Examiner during prosecution of the ’730 patent. *See, e.g.*, EX1004, at 919. The ’730 patent also lists McCloskey as one of its cited references. EX1001, 2.

115. McCloskey relates to “the use of barcodes and batch-stamps for *verifying the authenticity of PCR products and other sequence information.*” EX1012 (McCloskey), [0002]. McCloskey explains that PCR allows for “multiple copies of selected DNA sequences to be produced from limited amounts of DNA” but runs “the risk of amplifying contaminant DNA and can also result in a skewed yield of PCR products.” EX1012 (McCloskey), [0002]. Thus, McCloskey identifies “a need in the art for methods for verifying the authenticity of PCR products.” EX1012 (McCloskey), [0004].
116. McCloskey lists “examples of specific uses” for its invention including “analysis of limited amounts of template DNA for biomedical, ancient DNA, and forensic purposes.” EX1012 (McCloskey), [0006]. For example, McCloskey explains that “authentication of a forensic nucleic acid sample is crucial for comparison of samples from known and unknown individuals.” EX1012 (McCloskey), [0006]. Accordingly, McCloskey’s invention attempts to provide a “high level of confidence that a sample sequence represents a known sample processed on a certain day” such that “contaminating sequences can be identified by their molecular sequence tags, or lack thereof,” and differentiated from authentic sequences. EX1012 (McCloskey), [0006].

117. To address these issues, McCloskey's methods use "distinct sequence tags" to "authenticate a nucleic acid sequence with information, such as date of the experiment, and the sample identity." EX1012 (McCloskey), [0005]. For example, McCloskey's methods involve "bar-coding a nucleic acid molecule" by "contacting a target nucleic acid molecule in a sample with a bar-coded oligonucleotide." EX1012 (McCloskey), [0017].
118. McCloskey's "bar-coded oligonucleotide[s] comprise[] a first sequence complementary to the target nucleic acid molecule, a second sequence providing a random barcode, and a third sequence that is not complementary to any sequence in the sample." EX1012 (McCloskey), [0008]. Thus, each of McCloskey's bar-coded oligonucleotides is annealed to a target nucleic acid molecule using the "first sequence complementary to the target nucleic acid molecule" and then extended "using the target nucleic acid molecule as a template to produce a bar-coded target nucleic acid molecule." EX1012 (McCloskey), [0017]; *see also* EX1012 (McCloskey), [0029]–[0031].
119. McCloskey's methods further amplify the bar-coded target nucleic acid molecule. *See* EX1012 (McCloskey), [0009]. For example, McCloskey provides a method for "amplifying the bar-coded target nucleic acid

molecule using a primer that binds to the third sequence [of the bar-coded oligonucleotide] to product an amplification product.” EX1012 (McCloskey), [0032]; *see also* EX1012 (McCloskey), [0033].

120. “[B]y detecting the presence of the second sequence in the amplification product,” McCloskey authenticates the amplification product. EX1012 (McCloskey), [0034]. Therefore, the detection of the second sequence along “permits identification of valid sequences, and distinguishes valid sequences from contaminants and redundant sequences arising from template re-cloning.” EX1012 (McCloskey), [0005].

D. Bignell

121. Bignell (EX1013) is U.S. Patent No. 8,053,192 titled “Methods for Indexing Samples and Sequencing Multiple Polynucleotide Templates.” EX1013 (Bignell). Bignell “relates to methods for indexing samples during the sequencing of polynucleotide templates” so that “the source and sequence of each polynucleotide can be determined.” EX1013 (Bignell), 1:24-28. Bignell was filed on February 4, 2008.
122. Bignell acknowledged that “large-scale genotyping is demanding in terms of the cost of both materials and labour involved, and [] time ... especially if the methodology employed involves separate serial analysis of individual DNA samples.” EX1013 (Bignell) 1:51-55. Bignell further

recognized that pooling multiple samples together prior to sequencing “represents an effective technique for analysing large quantities of samples” but “[o]ne disadvantage of pooling samples prior to analysis is that information pertaining to individual DNA samples is lost” meaning “there is no easy method for discerning which individuals gave rise to a particular genotype.” EX1013 (Bignell), 1:66-2:3. Therefore, Bignell discloses methods where “the target sequence and the tag sequence can be determined” in order to identify the source of any particular sample in a multi-sample pool. EX1013 (Bignell), 3:54-59.

123. Therefore, Bignell’s invention provides methods to tag multiple DNA samples from different sources so as to identify the samples when sequencing them in a single reaction: “DNA from multiple sources can [] be tagged with different nucleic acid tags such that the source of the sample can be identified.” EX1013 (Bignell), 3:36-38 *See* EX1013 (Bignell), 3:36-50. Particularly, Bignell explains that its invention seeks to attach “the *same* tag to every molecule in the sample.” EX1013 (Bignell), 3:32-35.

124. Bignell’s methods involve pooling multiple tagged samples together and sequencing pooled samples. *See* EX1013 (Bignell), 4:19-29. Because each molecule from a particular sample was tagged with the same

sample-specific, but not molecule-specific, tag, each molecule can be traced back to its original sample. *See* EX1013 (Bignell), 3:36-50.

125. The use of sample-specific, but not molecule-specific, tags is critical to Bignell. In fact, Bignell considered and rejected a method that uses many unique sequences to tag different molecules within a single samples but explained that it “does not provide a method of analysis targets from multiple samples, but rather a method of sorting a mixture of amplified templates from a single sample.” EX1013 (Bignell), 3:24-27. Dr. Gabriel agrees that Bignell identified a “problem with enabling the method for individual samples rather than individual molecules.” EX1013 (Bignell), 3:28-35. Because “all 16,77,216 different sequence [tags] are obtained in a single mixture in the same tube,” ... “it does not permit attachment of the same tag to every molecule in the sample.” EX1013 (Bignell), 3:32-35. Thus Bignell, disclosed a method where DNA from multiple sources can be tagged with different nucleic acid tags and the source of the sample can be identified. EX1013 (Bignell), 3:36-38.

E. Chiang

126. Chiang (EX1033) is a publication titled “High-resolution mapping of copy-number alterations with massively parallel sequencing.” Chiang was published on November 30, 2008.

127. Chiang discusses issues and benefits that arise when identifying and mapping cancerous copy-number alterations using massively parallel sequencing as compared to Illumina sequencers. EX1033 (Chiang), 99. Chiang does not disclose tagging individual molecules with differentiator tags or collapsing molecules with an identical sequence into a single read count. Chiang instead uses a computational algorithm called SegSeq to map copy-number alterations determined from massively parallel sequencing data. EX1033 (Chiang), 100. Chiang tests its methodology using an Illumina sequencer, finding that that “copy-number estimates from massively parallel sequencing achieved greater sensitivity, higher dynamic range and greater precision for mapping breakpoints than similar estimates based on microarray hybridization.” EX1033 (Chiang), 101.

F. Bentley

128. Bentley (EX1017) is a publication titled “Whole-genome re-sequencing.” Bentley was published on October 18, 2006.

129. Bentley describes several new sequencing technologies, including microelectrophoresis, sequencing by hybridization, and sequencing by synthesis on arrays, which includes pyrosequencing (e.g., the Roche 454 system) and Solexa sequencers, which were the original version of

Illumina sequencers. EX1017 (Bentley), 546, 548. Bentley also describes the various cost, throughput, accuracy and completeness considerations for each of the types of novel sequencing technologies, and notes that “[t]he challenge for new technologies is to achieve massive improvement in one or more of these components without compromising the others.” EX1017 (Bentley), 548.

130. Bentley compares methods of analysis in existing technologies with those in new technologies. For example, Bentley explains that, while Sanger sequencing measures accuracy using an algorithm called Phred that estimates the probability of error for each base-call in a raw read, “new technologies work on a range of different principles that will require different ways to define good quality metrics. Success in this area is important to make reliable comparisons of the accuracy of different systems.” EX1017 (Bentley), 549. Bentley goes on to explain the challenges still existing in this area, noting that, “[a]t present, however, there are insufficient data available from most technologies. Base-by-base sequencing, as used in the Solexa system, provides information on the raw data collected at each base, enabling derivation of quality values that are analogous to Phred scores.” EX1017 (Bentley), 549. But Bentley continues: “More work is required in this area to assess the accuracy of

each sequencing method, to look for loss of accuracy in particular sequence contexts for each method, and to enable assembly of sequence data obtained using more than one technology.” EX1017 (Bentley), 549. In concluding, Bentley notes that the “next challenge will be to analyze all the data” collected from the new sequencing methods, once issues related to accuracy and completeness are addressed, but does not disclose any method for improving accuracy or addressing other factors still at issue, such as cost. EX1017 (Bentley), 549.

X. GROUNDS 1-2: THE CHALLENGED CLAIMS ARE NOT DISCLOSED BY CHEE

A. Petitioner And Dr. Gabriel Fail To Show That Chee Discloses A Method For “collapsing target:differentiator tag combinations observed more than once into a single count.”

131. Independent claim 1 of the ’730 patent requires “collapsing target:differentiator tag combinations observed more than once into a single count.” EX1001 (’730 Patent), 131:47-49. Dependent claim 2 of the ’730 patent further requires determining a consensus base call *after* the collapsing step has been performed. EX1001 (’730 Patent), 131:50-54 (“The method of claim 1, wherein said correcting step further comprises *collapsing* all sequence reads comprising a same differentiator tag into a single read *before determining a consensus base call* for the target nucleic acid molecule sequences.”). Dr. Gabriel does not show that

Chee discloses each of the “collapsing” step and the “determining a consensus base call” step because she improperly relies on the same disclosure and the same step in Chee—which describes only a consensus base call, *not* a collapsing step—for both separately claimed steps in the ’730 patent, as I describe further below. EX1005, ¶¶170-179, 184-186.

1. The ’730 Patent Discloses The “Collapsing” Step As A Limitation Separate And Distinct From “Determining A Consensus Base Call.”

132. A POSA would understand “collapsing” to have a separate and distinct meaning from “determining a consensus base call.” My understanding is based on the plain language in the claims, which is consistent with the specification, and Figure 5 of the ’730 Patent.
133. For example, dependent claim 2 of the ’730 patent requires determining a consensus base call *after* the collapsing step of claim 1 has been performed. EX1001 (’730 Patent), 131:50-54 (“The method of claim 1, wherein said correcting step further comprises *collapsing* all sequence reads comprising a same differentiator tag into a single read *before determining a consensus base call* for the target nucleic acid molecule sequences.”). This means that a consensus base call is performed on *already-collapsed* sequence reads. And for good reason – by performing the consensus base call on collapsed sequence reads to determine a

sample's genotype, amplification bias will be removed from consideration. Without collapsing, amplification bias errors (i.e., quantitative errors) would still be considered in the consensus base sequence.

134. The specification also describes “collapsing” as resolving issues related to skewing while using the “collapsed” sequences to derive a more accurate consensus base call for genotyping. EX1001 ('730 Patent), 65:2-14 (“Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10x average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed).”), EX1001 ('730 Patent), 28:3-9 (“Aspects of the disclosure are based, in part, on the discovery of methods for overcoming problems associated with systematic and

random errors (bias) in genome capture, amplification and sequencing methods, *namely high variability in the capture and amplification of nucleic acids and disproportionate representation of heterozygous alleles in sequencing libraries.*”).

135. Before determining a sample’s genotype at a given locus by performing a consensus base call, the method of the ’730 patent uses the “collapsing” step to ensure that any errors in amplification bias do not carry through and affect the subsequent analysis. EX1001, 72:63-73:5 (“For any target:differentiator tag combination observed more than once, all such reads are 'collapsed' into a single read before consensus basecalls are determined. This will cancel the effect of bias on consensus basecall accuracy.”) If the molecules with identical tag:combination sequences were **not** collapsed into one sequencing read before further analysis, the ’730 patent explains that considering sequence reads for all copies of a given original molecule could result in designating that locus with the wrong genotype. For example, the sample may be incorrectly designated as homozygous at a given locus instead of heterozygous. EX1001 (’730 Patent), 32:20-38 (“However, after multiple reads have been collected for a given locus, it is possible that due to bias (e.g., caused by PCR amplification steps), *a large fraction of the reads are derived from a*

single originating molecule. This would skew the population of target sequences observed, and would affect the outcome of the genotype call.

For example, it is possible that a locus that is heterozygous is called as homozygous, because there are only a few observations of the second allele out of many observations of that locus. . . . Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm *to significantly improve the accuracy and confidence of the genotype calls.*”).

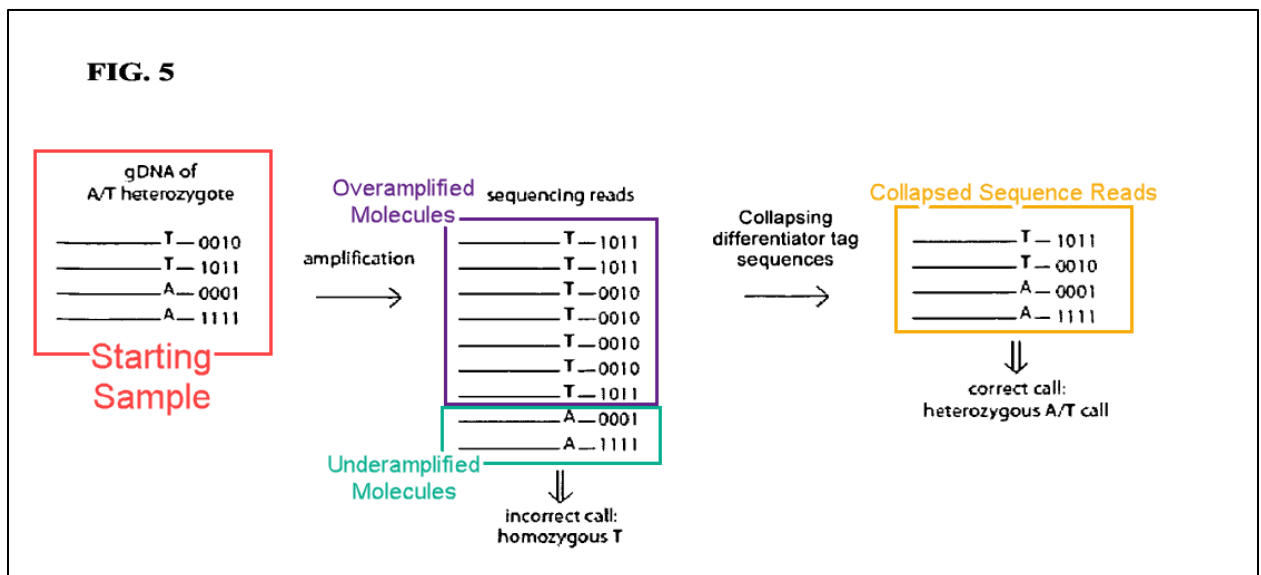
136. The '730 patent explains that “collapsing” a molecule containing the same target:differentiator tag combinations observed multiple times into a single count to correct for *skewing errors* (i.e., *quantitative errors*), which is different from reading multiple copies of the same original molecule to determine a consensus sequence to correct for *sequence errors*. EX1010 (Chee), 14:42-50, EX2024 (Gabriel Deposition), 46:2-25 (“Q. Okay. Absent a correct in supp later on, amplification bias would obscure quantification? A. It could. Q. Okay. And that would be a difference from, let's call it, like a sequencing error, by which I mean an error in the sequence read itself, right? A. *A sequence error is another type of error.* Q. Right. So in a sequence error, if you were -- you're

familiar with sequencing by synthesis? A. Yeah. Q. If you incorrectly incorporated the wrong nucleotide for a given position in the complementary strand during sequencing by synthesis, the sequence itself would have an error, right? A. Yes. Q. *And that's a different kind of error than the kind of quantitative error that, you know, would arise when amplification bias obscures the count?* A. *It's a different type of error."*

137. For example, the '730 patent describes "cancel[ling] the effect of *bias* on consensus base call accuracy" by collapsing any target:differentiator tag combination observed more than once into one read *before* consensus base calls are determined. EX1001 ('730 Patent), 72:63-73:5. This shows that collapsing is an additional step beyond determining a consensus base call, and for good reason. In the '730 Patent, collapsing corrects for *quantitative errors* arising from skewing during the amplification process to ensure disproportionately amplified molecules, for example, do not make it appear as though more copies of the original target nucleic acid molecule were present than there actually were if a consensus base call is later determined. If the collapsing step were not performed before determining the consensus base call for genotyping, the disproportionately amplified molecules may result in an incorrect

consensus base call or genotype, as shown by the '730 patent's Figure 5 below.

138. Figure 5 of the '730 Patent (reproduced below with annotations) shows the collapsing step, which is performed *independently* of downstream sequencing.



EX1001 ('730 Patent), Fig. 5 (annotations added).

139. The red box added to Figure 5 shows that, *before* amplification, the sample originally contained two copies of a certain heterozygous locus – that is, it contained two distinct nucleic acid molecules with the “T” allele (labeled with tags 0010 and 1011) and two distinct nucleic acid molecules with the “A” allele (labeled with tags 0001 and 1111). Those original nucleic acid molecules are amplified, but during amplification, skewing results, meaning that some of the nucleic acid molecules are

amplified more than others.

140. The purple and green boxes added to Figure 5 further show that, *after* amplification, there are four and three molecules with tags 0010 and 1011, respectively, and only one molecule each with tags 0001 and 1111. Amplification of those original molecules tagged 0010 and 1011 containing the “T” allele results in seven sequencing reads, while original molecules tagged with 0001 and 1111 containing the “A” allele result in only two sequencing reads. Thus, as Figure 5 shows, most of the sequenced reads will contain a “T” instead of an “A.”
141. If genotyping or a consensus base call were determined at this step, before collapsing, the genotype or base call determined would be incorrect. Specifically, because of the seemingly overwhelming number of sequencing reads attributed to the “T” allele, sequences with an “A” allele are assumed to be random errors arising during the amplification or sequencing workflow instead of a true difference in the sequence of the original molecules. As a result, genotyping results, for example, in incorrectly designating the sample as homozygous at that locus (i.e., contains two “T” alleles) instead of heterozygous (i.e., contains a “T” allele and an “A” allele).
142. On the other hand, as the yellow box added to Figure 5 shows, by using

the invention, collapsing the sequencing reads that have the same target:differentiator tag combination into one read will allow recognition of other copies of the same target sequence as arising from different original molecules (i.e., additional copies of the target sequence) with the “A” version of the allele instead of sequencing errors. *See also* EX1001 (’730 Patent), 73:3-5. Thus, after collapsing, the sample can be correctly determined to have a heterozygous A/T genotype, since the analysis will be performed by considering only one sequence read for a single original molecule, not all amplicons that arose from that molecule. EX1001 (’730 Patent), 32:29-34 (“However, if information is available on differentiator tag sequences, this situation could be averted, because the over-represented allele would be seen to also have an over-represented differentiator tag sequence (i.e., the sequences with the overrepresented differentiator tag sequence all originated from the same single molecule). Therefore, the sequences and corresponding distribution of differentiator tag sequences can be used as an additional input to the genotype-calling algorithm to significantly improve the accuracy and confidence of the genotype calls.”).

143. Further, the ’730 Patent explains that “collapsing” by itself allows target nucleic acid molecules to be mapped to the correct place on the genome.

For example, Example 2 describes “collapsing” before determining a consensus base call as a method for detecting quantitation biases. EX1001 (’730 Patent), 72:48-49. It further notes that any count for a target:differentiator tag greater than one should be “collapsed” into a single read because “any particular target:differentiator tag mapping will not occur more than once by chance, and therefore will only be observed if bias was present somewhere in the sample preparation process.” EX1001 (’730 Patent), 72:63-73:2. Collapsing thus determines the number of places within a sample genome that a given sequence associated with a differentiator tag appears. A POSA would have understood this to aid in mapping approximate sequences when aligned with a reference genome. On a similar note, the ’730 Patent explains that collapsing can be used to detect the presence of an insertion or deletion, which does not require determining the sequence of a locus but rather considers the number of original copies of a given sequence (e.g., repetitive region) present. EX1001 (’730 Patent), 56:64-57:3 (“In contrast, in some embodiments, aspects of the invention relate to detecting the presence of an insertion or deletion at a genomic locus without requiring the locus to be sequenced (or without requiring the entire locus to be, sequenced). Aspects of the invention are particularly

useful for detecting an insertion or deletion in a nucleic acid region that contains high levels of sequence repeats.”).

144. Therefore, in my opinion, the ’730 Patent provides clear guidance that would allow a POSA to understand the difference between “collapsing” sequence reads with the same sequence and using those collapsed reads to determine a consensus base sequence, making the “collapsing” step of the ’730 Patent different from the consensus base sequence determined by Chee.

2. Chee Does Not Disclose The Collapsing Step.

145. Petitioner and Dr. Gabriel argue that Chee discloses the collapsing step as described in the ’730 Patent because it counts multiple reads with the same unique identifier as “one instance of the ROI.” EX1005 (Gabriel Decl.), ¶173, Pet. at 32.
146. I disagree that Chee discloses collapsing, because Dr. Gabriel and Petitioner incorrectly conflate the collapsing step with determining a consensus base call for remedying sequence errors. Dr. Gabriel cites to the same single-step disclosure in Chee for *both* collapsing and determining a consensus base call. EX1005 (Gabriel Decl.), ¶¶175-178, ¶185. For example, as described further above, claim 1 of the ’730 patent requires correcting for errors or bias “by collapsing.” EX1001 (’730

Patent), 131:47-49. Dr. Gabriel cites Chee's disclosure that Chee reads multiple sequences arising from the same construct and compares them to attain a consensus base sequence, which allegedly corrects for errors (EX1005 (Gabriel Decl.), ¶177 (citing EX1010 (Chee), 14:42-50), but this process describes "determining a consensus base call," which corrects for sequence errors, not "collapsing." Tellingly, Dr. Gabriel goes on to explain the exact same disclosure from Chee in the context of determining a consensus base call. EX1005 (Gabriel Decl.), ¶178.

147. To be clear, Chee only recites reading multiple sequences with the same *sample identifier*. EX1010 (Chee), 14:42-50. In my opinion, a POSITA would not understand collapsing to involve Chee's method of reading multiple sequences of the same construct to correct for error or bias by "collapsing" for at least two reasons. First, Chee only mentions the possibility of using its methods to reduce amplification bias, but points to no actual method of "collapsing" to do so, and its method of determining a consensus base sequence to detect sequence errors does not resolve quantitative errors when reading a given sequence as collapsing does. Second, Chee's express disclosure which requires reading multiple clonal copies of the same molecule is the exact opposite of "collapsing" molecules with the same differentiator tag.

i. Chee’s Generic Disclosures About Skewing And Amplification Bias Do Not Disclose A Method For “Collapsing” As Required By The ’730 Patent.

148. Petitioner and Dr. Gabriel do not show that Chee discloses a method for “collapsing” multiple copies of a single construct into one count simply by pointing to generic disclosures in Chee that say its method can be used to reduce amplification bias.
149. Dr. Gabriel supports her analysis by opining that the ’730 patent shows the collapsing step “by counting the number of unique tags associated with each sequence (and ignoring duplicates) to determine the number of times a sequence appeared in the original sample.” EX1005 (Gabriel Decl.), ¶175. In support of her conclusion, Dr. Gabriel points to statements in the ’730 patent that disclose collapsing identical target:differentiator tag sequence combinations to improve accuracy when determining a sample’s genotype (EX1001 (’730 Patent), 30:48-52, 72:63-73:5), but neither of these statements, when read in context, explain “collapsing” by determining a consensus base sequence (EX1005 (Gabriel Decl.), ¶175).
150. Instead, the collapsing step is a specific process performed prior to any subsequent analysis so that multiple identical sequence reads derived

from the same original molecule are not considered more than once in a genotype call. By “collapsing” *first prior to any consensus sequence analysis*, read redundancy will be eliminated so that sequencing reads are statistically represented correctly. Skewed amplification, for example, of certain alleles over others will no longer improperly affect the ultimate genotype determination, since any disproportionately amplified molecules will not be considered in determining the consensus base call. Thus, collapsing is not simply obtaining a single consensus base call for all reads with the same unique tag and using that for subsequent analyses, because doing so would not eliminate amplification bias.

151. Chee’s generic disclosures used by Dr. Gabriel to support her incorrect reading of the ’730 patent do not disclose the more nuanced meaning of “collapsing” that the ’730 patent provides. For example, Dr. Gabriel points to broad language to suggest that Chee’s methods could be used to identify amplification bias and infer the number of starting molecules of the initial ROI present in a sample. EX1005 (Gabriel Decl.), ¶¶174-175 (citing, *e.g.*, EX1010 (Chee), 9:5-9, 24:1-8, 10:41-46, 23:50-64). Dr. Gabriel also points to generic language in Chee about identifying “technical bias” introduced during amplification that can “skew” the amplicons in the final sample, and using the unique identifiers to

recognize constructs that are “identical by descent” to supposedly correct for amplification bias. EX1005 (Gabriel Decl.), ¶¶172-173 (citing, *e.g.*, EX1010 (Chee), 5:27-36, 9:18-21, 10:36-46). But Dr. Gabriel does not cite to any part of Chee that explains *how* this skew is reduced (i.e., how Chee collapses) if Chee only achieves a single sequence read by determining the consensus base sequence of multiple clonal copies of the same molecule.

152. These generic citations actually support Dr. Gabriel’s analysis of the process for determining a consensus base sequence, but, as I further describe above, this process is not the same as remedying amplification bias errors through collapsing, which is a separate and distinct concept in the ’730 Patent. *See, e.g.*, EX1005 (Gabriel Decl.), ¶¶177-178. For example, Chee’s disclosure that uniquely associating identifiers with individual molecules in the starting sample to later allow amplification products to be identified as “identical by descent” (EX1010 (Chee), 5:27-36) is used by Chee to read *multiple* clonal copies of the same construct to determine an *accurate consensus base sequence*, not for *collapsing* identical sequence reads into *one* for subsequent quantitative analysis (EX1010 (Chee), 14:46-50 (“Thus, a variation in the sequence of a single copy of a construct is most likely to be a synthesis or sequencing error if

it differs from a consensus sequence derived from multiple clonal copies of the construct, *which are identified as identical by descent via the shared unique identifier.*”)).

153. Chee also broadly discloses that associating the unique identifier with each construct can be used to determine the number of starting molecules of each initial ROI present (EX1010 (Chee), 10:41-46), but this again does not explain *how* to reduce *skewing errors* through collapsing identical sequence reads into one prior to subsequent analysis. Instead, it only explains that the number of starting molecules can be used to reduce *sequence errors* to determine the entire accurate sequence of all molecules present in a sample. EX1010 (Chee), 10:32-41 (“The sets of molecules of the invention will contain some ROI sequences that are exactly those designed or isolated from a sample, and some that have errors introduced in the subsequent manipulation of Such nucleic acids. The use of the sets of constructs and the methods disclosed herein allow one skilled in the art to identify the correct sequences within a master set or a selected Subset of a masterset, and can identify correct ROIs from ROIs that contain errors due to technical realities of chemical synthesis, amplification, and the like.”). But reducing sequence errors by determining the most probable nucleotide at each position of an ROI will

not remove quantitative errors from that consensus base call analysis.

154. Furthermore, Dr. Gabriel says that Chee collapses because, “where it reads multiple sequences with the same unique identifier, it counts those as one instance of the ROI (not multiple instances).” I disagree with this characterization of Chee. EX1005 (Gabriel Decl.), ¶173.
155. First, in support of this statement, Dr. Gabriel again only cites to generic language in Chee to say association of an ROI with a unique identifier can be used to determine the starting number of molecules, and that this corrects for over-represented or under-represented constructs from amplification bias. EX1005 (Gabriel Decl.), ¶174. But these parts of Chee do not say anything about counting an ROI (or a construct containing an ROI) only once, nor do they disclose any precise *method* for actually using Chee for quantification corrections arising from amplification biases.
156. Second, simply counting sequences with the same unique identifier as arising from only one ROI once a consensus base call has been determined again does not describe the “collapsing” method in the ’730 Patent. Assuming Dr. Gabriel means that multiple sequences with the same unique identifier will count as one construct (not one ROI), the ’730 Patent does something more, which is to ensure that any “skewing”

resulting from amplification bias does not affect subsequent analyses (e.g., genotyping *or* consensus base calls). As I describe further below, Chee does not draw this connection and instead reads every clonal copy of an original construct without regard to whether amplification bias affected the resulting number of amplicons of that construct. Simply “counting” a sequence only once based on this consensus sequence determined does not solve one of the key problems identified and solved by the ’730 Patent, which was to use molecular identifiers to correct for amplification bias.

157. For at least these reasons, Dr. Gabriel fails to point to *any* part of Chee that actually discloses a *method* for “collapsing” constructs observed more than once into a single count for the purpose of eliminating errors resulting from skewing during amplification.

ii. Chee’s Disclosures Show Reading Multiple Clonal Copies Of The Same Molecule, Which Is The Opposite Of “Collapsing . . . Into A Single Count.”

158. Dr. Gabriel’s interpretation of Chee does not appreciate that the ’730 Patent teaches performing “collapsing” by considering only *one* sequence read of a molecule per unique target:differentiator tag combination, *not* multiple clonal copies of molecules. I agree with Dr.

Gabriel that Chee discloses identifying and sequencing multiple copies of the same original molecule and comparing them to determine a *consensus* sequence. EX1005 (Gabriel Decl.), ¶177; EX1010 (Chee), 14:42-50 (“*Multiple instances* of the *same molecule* can be sequenced Thus, a variation in the sequence of a *single copy* of a construct is most likely to be a synthesis or sequencing error if it differs from a consensus sequence derived from *multiple clonal copies* of the construct, *which are identified as identical by descent via the shared unique identifier.*”). But this exact point supports my conclusion: Chee reads *every amplicon* arising from the same original molecule to get a precise sequence. It does *not* first collapse amplicons with identical sequences into one sequence read correlating to one original construct and unique differentiator tag prior to subsequent analyses (e.g., performing a consensus base call), as the yellow box added to Figure 5 above shows. Nor does Chee align the *collapsed* sequence reads arising from other original constructs (i.e., copies) of the same ROI to determine the genotype of that ROI. This point alone shows that Chee does not disclose “collapsing” sequences arising from the same original molecule into one read.

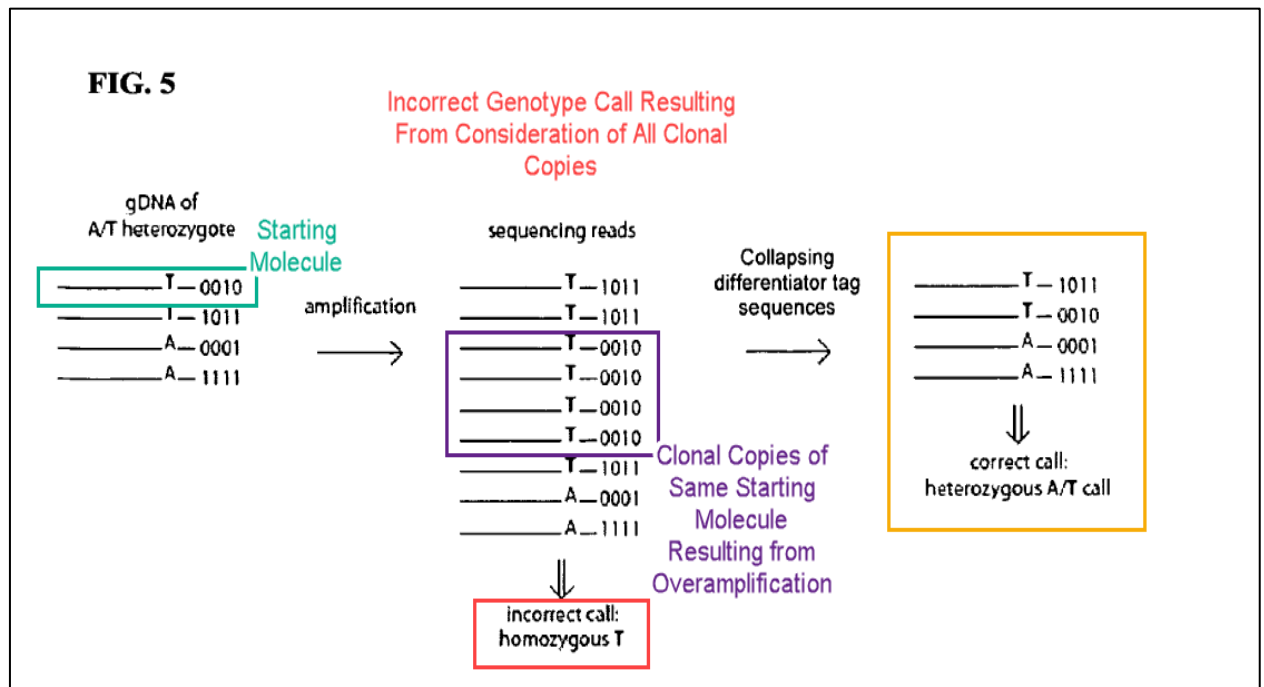
159. Dr. Gabriel only points to parts of Chee that show deriving a consensus

base sequence by reading multiple clonal copies of the same construct and, as support, points to Figure 2 of Bentley (EX1017), a 2006 article that allegedly shows “[i]t was well-known to those in the art that deriving a consensus sequence involved considering sequencing reads from multiple overlapping copies of a DNA sequence to come up with a single sequence for the DNA segment.” EX1005 (Gabriel Decl.), ¶¶176-178. But deriving a consensus base sequence by reading the same position of an original molecule multiple times using *clonal* copies of the same construct is *not* the same as first collapsing identical molecules into one sequence read and using those *collapsed* versions representing multiple copies of a locus existing in the sample to then derive a genotype. This is because including sequence reads from clonal copies of the same starting molecule in the consensus base sequence determined could still result in assigning the incorrect genotype to the sample due to skewed amplification, for example, of certain alleles over others, causing an incorrect perception of the frequency by which those alleles occur in the sample.

160. In fact, Figure 2 of Bentley, which Dr. Gabriel points to as supposed evidence of correcting errors during amplification by using the “collapsing” method of the ’730 patent, is actually an example of a

method that does **not** use **any** form of molecular tagging, much less does it use molecular tags to **collapse**. Ultimately, similar to Chee, the methods described in Bentley align with the method disclosed in the purple box in Figure 5 of the '730 patent (reproduced below with annotations), which is expressly disclosed as **not** performing the method of the '730 patent and that instead **causes** the issues solved by the '730 patent's disclosed method.

161. The purple box shown below in Figure 5 of the '730 patent, Chee, and Bentley all describe reading multiple clonal copies of the same starting molecule before subsequent analyses (e.g., genotyping). See, e.g., EX1010 (Chee), 14:42-46 ("Multiple instances of the same molecule can be sequenced, and the ROI of an individual parent construct can be identified by its unique identifier, as molecules with identical identifiers arise via amplification of the same original construct."), EX1017 (Bentley), 546 ("The use of clonal single-molecule templates is very useful for studying heterogeneous samples; for example, either for detecting heterozygous SNP (single nucleotide polymorphism) alleles in diploid DNA or, alternatively, for identifying somatic mutations in a tumour. Because each sequencing template is clonal in origin, high-quality base-calls are generated separately for each variant.").



EX1001 ('730 Patent), Fig. 5 (annotations added).

162. However, as the yellow box added to Figure 5 of the '730 patent shows, reading multiple clonal copies of the same sequence read is *not* the same as collapsing regions with identical sequence reads prior to subsequent analyses (i.e., genotyping or determining a consensus base sequence) such that each identical sequence read is considered only once. Neither Chee nor Bentley fill the existing gaps to show that a POSA would have had the requisite knowledge to collapse sequences with identical target:differentiator tag sequences, in the manner shown, prior to genotyping or determining a consensus base sequence to reduce errors resulting from skewing, for example.

163. For at least these reasons, in my opinion, Chee does not disclose a method

for “collapsing” copies of the same or substantially the molecule into a single count as described by the ’730 Patent, nor does it disclose a method for correcting for amplification bias. Therefore, it is my opinion that Chee does not anticipate the ’730 Patent.

B. Petitioner And Dr. Gabriel Fail To Show That Chee’s Provisional Application Discloses A Method For “obtaining nucleic acids from a biological sample.”

164. Chee’s provisional application, which I understand is dated April 30, 2009, does not make any mention of performing its method on *naturally-occurring* nucleic acids. EX1011 (Chee Provisional), [00029], [00060], [00066], [00069]; Pet. at 27. And Dr. Gabriel does not even make reference to the “obtaining a biological sample” limitation found in claims 1 and 9 of the ’730 Patent in her analysis of Chee’s provisional application, nor does she point to a place in Chee’s provisional application where this limitation can be found. EX1005 (Gabriel Decl.), §VII.A.2.(a); *see also id.*, ¶¶137-143. Dr. Gabriel instead only makes reference to adding unique identifiers to “a sample of isolated nucleic acids containing the ROI,” but does not point to any part of Chee’s provisional application to explain where those nucleic acids are obtained from. EX1005, ¶123.

165. I would understand the ’730 Patent to be directed toward performing its

method on naturally-occurring nucleic acids. As I describe further in Section XI below, claims 4 through 8 all require performing the method on naturally-occurring biological samples, such as blood, biopsy tissues, circulating tumor molecules, samples collected from a patient, and maternally-circulating nucleic acid molecules. EX1001, 131:59-38. The specification also states that invention is used “for determining genotypes in *patient samples*” (EX1001, 1:23-24), fulfills the need “for *reliable high throughput analysis of many genetic loci in large numbers of patients*” (EX1001, 1:42-53), and “may be used in connection with *any medical evaluation* where the presence of one or more alleles at a genetic locus of interest is relevant to a *medical determination* (e.g., risk or detection of disease, disease prognosis, therapy selection, therapy monitoring, etc.),” EX1001, 19:8-13.

166. However, I would understand Chee’s provisional application to be directed to a method for synthetic biology without disclosing performing its inventions on naturally-occurring DNA. For example, Chee’s provisional application explains that its invention “relates to the field of oligonucleotide selection and *synthetic biology*.” EX1011, [0001], [0003]-[0005]. It further explains that it is “*primarily described in the context of synthetic biology*” without explaining any method for creating

constructs containing ROIs from a naturally-occurring sample. EX1011, [0066]. Further, Chee's provisional application explains that its methods "can be used to assemble oligos into larger constructs, and these can be used to assemble larger nucleic acids using the isolated ROIs" and that "[b]y obtaining long, highly pure oligos of desired sequence, it will be possible to assemble relatively long constructs by annealing the oligos together so that hybridization ensures that correct sequences are brought together." EX1011, [0129], [0130].

167. Although the '442 Provisional makes reference to analyzing "nucleic acids derived from a biological sample," (EX1011, [0060], [0069]) the '442 Provisional does *not* describe the biological sample as being naturally-occurring. Read in the context of the rest of the '442 Provisional, I would understand this language to disclose analyzing nucleic acids derived from a *synthetically-created* biological sample, not analyzing them obtained from a naturally-occurring sample. EX2019 (Keasling), 65 ("Synthetic biology is the design and construction of *new biological components*, such as enzymes, genetic circuits, and cells, or the redesign of existing biological systems.").
168. Taken in context, I would read Chee's provisional application as disclosing a method directed to synthetically-created biological

molecules, not naturally-occurring biological molecules, such as those collected from patient samples.

169. For these reasons, I do not understand Chee's April 30, 2009, provisional application as disclosing performance of its invention on naturally-occurring biological samples.

**XI. GROUND 3: THE CHALLENGED CLAIMS ARE PATENABLE
OVER CHEE IN VIEW OF GAHAN**

170. As described above in Section X, Chee does not disclose the "collapsing" step as described in the '730 patent. Petitioner and Dr. Gabriel do not point to any part of Gahan that would allegedly further disclose the "collapsing" step. EX1005 (Gabriel Decl.), ¶¶222-232. And Petitioner and Dr. Gabriel cannot rely on Gahan for any such support because Gahan makes no mention of next-generation sequencing or tagging, either on a sample level or on a molecular level. Therefore, because the "collapsing" step of claim 1 is not disclosed by either Chee or Gahan, the challenged dependent claims are at least each patentable over Chee in view of Gahan.

171. Instead, Petitioner and Dr. Gabriel allege that Gahan discloses limitations relating to the source of target nucleic acid molecules found in dependent claims 5, 6, and 8, which all depend on independent claim 1. Dependent

claim 5 further recites, “wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.” EX1001 (’730 Patent), 132:27-29. Dependent claim 6 further recites, “wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.” EX1001 (’730 Patent), 132:30-32. Dependent claim 8 further recites, “wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.” EX1001 (’730 Patent), 132:35-37. However, I disagree with Dr. Gabriel and Petitioner’s conclusion that a POSA would have been motivated to use Chee’s methods in view of Gahan to screen viral, bacterial, tumor, or fetal DNA.

172. First, Chee and Gahan are in two different fields. As I describe above in Section X.B., the ’730 Patent is directed to performing its methods on *naturally-occurring* DNA. But Chee primarily focuses its methods on *synthetically-created* DNA. For example, Chee explains that its method “allows the use of very small numbers of individual molecules to make desired larger synthetic constructs.” EX1010 (Chee), 11:39-41; *see also id.*, 12:1-6 (“For example, when the ROIs of the constructs are envisioned to be used in synthetic biology for construction of larger molecules, the construct can be synthesized as a single unit comprising the ROI, the

unique identifier, and any other desired elements that may be necessary for further analysis, amplification, excision, and/or use of the constructs.”). Chee also provides a method for excising ROI sequences determined to be accurate from using its invention to determine the sequence of a larger region of the genome. EX1010 (Chee), 20:32-49 (“In FIG. 2, constructs 201 and 203 are general schematics of single-stranded constructs comprising: two amplification sites, one found substantially at the 5' end of the oligonucleotide and the other provided substantially at the 3' end of the oligonucleotide; a unique identifier that distinguishes each initial construct from the other initial constructs in the master set; and excision sites that allow the separation of the ROI from the remaining element of the construct, thus enabling separation and isolation of the ROI for further use in various synthetic biology, quantitative measurement, or other methodologies.”). Although Chee notes the possibility of performing its method on naturally-occurring DNA or RNA fragments “isolated from a sample,” Chee does not provide any specific reason that would lead a POSA to explain its method could or should be used for diagnostic purposes on cell-free DNA. EX1010 (Chee), 11:49-56. And despite Chee’s focus on resolving issues related to next-generation sequencing, Gahan mentions nothing of the sort.

173. Further, as I also describe above, Chee's provisional application, which I understand is dated April 30, 2009, does not make any mention of performing its method on naturally-occurring nucleic acids. EX1011 (Chee Provisional), [00029], [00060], [00066], [00069]; Pet. at 27. And Dr. Gabriel does not even make reference to the "obtaining a biological sample" limitation in her analysis of Chee's provisional application. EX1005, §VII.A.2.(a). Dr. Gabriel only makes reference to adding unique identifiers to "a sample of isolated nucleic acids containing the ROI," but does not point to any part of Chee's provisional application to explain where those nucleic acids are obtained from. EX1005, ¶123. Thus, Chee's provisional application, purely focused on synthetic biology, does not even provide any motivation or reason to combine it with a reference directed to naturally-occurring nucleic acids.

174. Second, Gahan explains that the field of cell-free DNA was just developing at the time of the invention. For example, as Gahan explains, *"[w]hile some workers have made an attempt to determine the kit with the greatest yield* and the largest range of fragment sizes for both RNA and DNA, *there is still a lack of clarity* concerning the user of the broad range of kits and their applicability." EX1015 (Gahan), 3. Gahan also notes that, *"the methodology is far from consistent*, and if there are to

be reliable protocols for diagnostic purposes, there needs to be a form of quality assurance for these procedures.” EX1015 (Gahan), 3. But Chee notes that, for its method to be performed properly, there needed to be enough constructs “to ensure a high probability of obtaining at least one correct sequence or to confirm a correct sequence associated with a particular unique identifier.” EX1010, 2:44-51. A POSA knowing they had to have a large enough population of *synthetically-created* starting nucleic acid material to accurately perform Chee’s method would not look to a paper describing inconsistencies on how to isolate a substantial amount of cell-free DNA to even perform the method.

175. Dr. Gabriel and Petitioner ignore these disclosures of uncertainty in Gahan, instead pointing to Chee’s blanket incorporation-by-reference of the *Genetic Variation Manual* as reason for combining Chee’s method with Gahan, despite Chee’s overwhelming focus on performing its methods on synthetically created nucleic acid molecules for synthetic biology. Pet., 41, EX1005 (Gabriel Decl.), ¶222. But the *Genetic Variation Manual* referenced in Chee is 472 pages and contains chapters ranging from study design to data analysis using a variety of techniques, including software algorithms. EX2022, *Genetic Variation Manual* Table of Contents. And Chee’s brief mention of it in a citation alongside

twelve other lengthy textbooks or laboratory manuals would not point a POSA to specific disclosures regarding isolation and analysis of cell-free DNA generally, much less bacterial, viral, tumor, and fetal DNA specifically. EX1010 (Chee), 8:17-37. Dr. Gabriel and Petitioner fail to explain why a POSA reading Chee would look to Gahan, which expressed uncertainties surrounding current cell-free DNA isolation, and specific closures made in this specific laboratory manual, specifically given that Chee is a method primarily designed for *synthetic* biology analyses, not naturally-occurring nucleic acids. For example, since the method of Chee was primarily designed to address the high degree of errors that arise when synthetically creating nucleic acid molecules, I would not expect a POSA to refer to a paper directed to a different problem (i.e., isolating naturally-occurring nucleic acids from biological samples) that would introduce greater uncertainty.

176. For at least these reasons, in my opinion, Dr. Gabriel and Petitioner do not provide any motivation to combine Chee in view of Gahan for Petitioner's Ground 3.

**XII. GROUND 4: THE CHALLENGED CLAIMS ARE PATENTABLE
OVER MCCLOSKEY IN VIEW OF BIGNELL**

**A. McCloskey And Bignell Do Not Disclose Or Render Obvious The
Limitations of Claim 1**

177. Claim 1 of the '730 patent requires “correcting for error or bias introduced during said workflow by *collapsing target:differentiator tag combinations* observed more than once into a single count.” EX1001, 131:47-49. A POSA would thus understand that claim 1 of the '730 patent requires correcting for errors by using target nucleic acids and differentiator tags as a combination. EX1001, 131:47-49.
178. Dr. Gabriel states that “Bignell and McCloskey together teach” the correcting for error or bias limitation but Dr. Gabriel only relies on McCloskey as support for this limitation. EX1005 (Gabriel Decl.), ¶¶ 291-295. Specifically, Dr. Gabriel asserts that “McCloskey ... teaches ‘collapsing’ duplicate ‘target:differentiator tag combinations ... into a single count.’” EX1005 (Gabriel Decl.), ¶ 295. Dr. Gabriel misreads McCloskey.
179. A POSA would have understood that McCloskey discloses methods that detect the *barcode alone*, or the barcode with a batch-stamp, to correct for skewing that can arise from PCR amplification. EX1012 (McCloskey), [0009], [0010], [0034], [0035]. McCloskey provides

methods for “verifying the authenticity of PCR products and other sequencing information.” EX1012 (McCloskey), [0002], [0004]. In order to verify the authenticity of PCR products, McCloskey incorporates “the use of barcodes and batch stamps.” EX1012 (McCloskey), [0002], [0008]. McCloskey’s barcoded-oligonucleotides require “a second sequence providing a random barcode” that are annealed to a “target nucleic acid molecule.” EX1012 (McCloskey), [0009]. McCloskey’s methods disclose “authenticating the amplification product by detecting the presence of *the second sequence* [providing the random barcode] in the amplification product.” EX1012 (McCloskey), [0009], [0034]. Further embodiments in McCloskey disclose introducing a batch-stamp with experiment information and authenticating PCR products using barcodes and batch-stamps, but none of McCloskey’s embodiments require the use of the target nucleic acid molecule to authenticate or correct for any errors. EX1012 (McCloskey), [0010], [0035]. In fact, there is no disclosure in McCloskey whatsoever describing the use of barcodes or batch-stamps *in combination with* the target nucleic acid molecule to correct for errors or bias by collapsing those combinations into a single count.

180. A POSA would have understood that McCloskey does not disclose

“correcting for errors or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count” at least because its methods do not use the target nucleic acid sequence to correct for errors or bias. EX1012 (McCloskey), [0009], [0010], [0034], [0035]; EX1001, 131:47-49. Thus, Dr. Gabriel’s assertion that McCloskey “teaches ‘collapsing’ duplicate ‘target:differentiator tag combinations ... into a single count’ is incorrect. EX1005 (Gabriel Decl.) ¶ 295.

181. Dr. Gabriel suggests that the collapsing step in the ’730 patent is the same as the authentication step in McCloskey. EX1005 (Gabriel Decl.) ¶ 295 (citing EX 1001, 28:32-44, 30:48-52, 72:63-73:2). Dr. Gabriel’s suggestion is incorrect. The methods of McCloskey sought to “verify[] the authenticity of PCR products” by detecting a barcode sequence, alone or in combination with a batch-stamp, to ensure that the sequence being analyzed is an authentic sequence and not a “contaminant” or “redundant” sequence. EX1005 (McCloskey), 0002], [0005], [0009]. A POSA would have understood that McCloskey’s methods do not correct for errors by collapsing target:differentiator tag combinations because its methods *only detect a barcode sequence*, or barcode and batch-stamp, to *authenticate the PCR product*. EX1012 (McCloskey), [0009], [0010,

[0034], [0035].

182. A POSA would have understood that collapsing target:differentiator combinations to correct for error or bias is different than authenticating an amplification product by detecting a barcode sequence. For example, figure 5 depicts a method for making diploid genotype calls in which repeat target:differentiator tag combination[s] are collapsed” and the “effect of bias on consensus basecall accuracy” is canceled. EX1001, 73:2-5, Figure 5; *see also* EX1005, (Gabriel Decl.) ¶ 295 (citing EX 1001, 28:32-44 (“[A] threshold number of independently isolated molecules (e.g., unique ***combinations of differentiator tag and target sequences***) is analyzed to determine the genotype of a subject.”), 30:48-52 (“FIG. 5 depicts non-limiting method for genotyping based on ***target and differentiator tag sequences***. Sequencing reads of ***target and differentiator tags sequences are collapsed*** to make diploid genotype calls.”), 72:63-73:2 (“For any ***target:differentiator tag combination*** observed more than once, ***all such reads are ‘collapsed’ into a single read*** before consensus basecalls are determined.”)). Unlike the methods to authenticate an amplification product in McCloskey, the ’730 patent explains that using the sequence of the target nucleic acid molecule and a differentiator tag to collapse target:differentiator tag combinations

provide methods to accurately make genotype calls and subsequently increase the accuracy when determining consensus base calls. EX 1001, 28:32-44, 30:48-52, 72:63-73:5, Figure 5. A POSA would have understood that McCloskey's authenticating step does not use target nucleic acids and thus does not provide methods to make genotype calls or determine consensus base calls. Without the target nucleic molecule sequence, a POSA would not be able to make genotype calls or determine consensus base calls. EX 1001, 28:32-44, 30:48-52, 72:63-73:5, Figure 5.

183. Dr. Gabriel claims that "Bignell and McCloskey together teach" "correcting for error or bias during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count" but Dr. Gabriel does not cite to Bignell for any support. EX1005 (Gabriel Decl.), ¶¶ 291-295. **First**, Dr. Gabriel focuses solely on McCloskey to teach the correcting step. EX1005 (Gabriel Decl.), ¶¶ 291-295. But as discussed above, McCloskey discloses detecting a barcode to authenticate PCR products and does not collapse target:differentiator tag combinations to correct for errors or bias. **Second**, Bignell does not fix McCloskey's flaws. Dr. Gabriel does not cite to any portion of Bignell that would equate the authenticating step in McCloskey to the correcting

for error step in the '730 patent. EX1005 (Gabriel Decl.), ¶¶ 295. In fact, Dr. Gabriel does not cite to Bignell for this limitation at all. EX1005 (Gabriel Decl.), ¶¶ 291-295. That is because Bignell is focused on sample tagging and does not tag individual target nucleic acid molecules with different differentiator tags. *See supra* Section IX.D; EX1013 (Bignell), 1:24-32. Instead, Bignell teaches tagging each molecule from a sample with the **same tag** so that “the source and sequence of each polynucleotide can be determined.” EX1013 (Bignell), 1:24-32. Furthermore, Bignell does not correct for errors or bias by collapsing “target:differentiator tag combinations into a single count” because Bignell does not collapse any sequences. A POSA would have understood Bignell focused on Thus, Bignell proposed a different solution to problems with multiplexing multiple samples but did not focus on error correction within nucleic acid molecules in a single sample. Therefore, a POSA would have understood that Bignell does not disclose “correcting for error or bias introduced during said workflow by **collapsing target:differentiator tag combinations** observed more than once into a single count.” EX1001, 131:47-49.

B. A POSA Would Not Have Been Motivated To Combine Bignell and McCloskey

1. McCloskey And Bignell Teach Away From Each Other.

184. Dr. Gabriel's analysis of the motivation to combine McCloskey and Bignell is flawed because McCloskey's and Bignell's goals and methods for adding tags to molecules are antithetical to each other and teach away from each other. Particularly, while McCloskey attaches a unique tag to every molecule in a small sample set in Sanger sequencing (EX1012 (McCloskey), [0043]-[0104]), Bignell teaches attaching the *same* tag to every molecule in a massive sample set in next-generation sequencing. EX1013 (Bignell), 3:32-35. Indeed, in Bignell's discussion of the background of the invention, Bignell considers and rejects the concept of applying a different tag to each molecule because it does not allow for attaching the same tag to every molecule in a sample set comprising millions of sequences:

The problem with enabling the method for individual samples rather than individual molecules is that the tags are synthesised in a combinatorial manner, meaning that all 16,777,216 different sequences are obtained in a single mixture in the same tube. *Whilst this is ideal for treating one sample such that each individual molecule in the sample carries a different tag, it does not permit attachment of the same tag to every molecule in the sample.*

EX1013 (Bignell), 3:28-35.

185. Dr. Gabriel never addresses this disclosure by Bignell in her discussion of the supposed motivation to combine McCloskey and Bignell. But a POSA would not have been motivated to combine McCloskey's method for applying a different tag to every sequence in a small sample set with Bignell's method when Bignell actually argues against such an application.

2. It Would Have Been Cost-Prohibitive To Combine McCloskey And Bignell Because They Are Directed To Different Sequencing Technologies.

186. McCloskey provides methods to "authenticate a nucleic acid sequence with information, such as the date of the experiment, and the sample identity." EX1012 (McCloskey), [0005]. McCloskey states that its invention "permits identification of valid sequences, and distinguishes the valid sequences from contaminants and redundant sequences arising from re-cloning." EX1012 (McCloskey), [0005].

187. McCloskey's invention is in the context of Sanger sequencing, a predecessor to next-gen sequencing. Section IX.C.. While McCloskey "does not describe techniques for the sequencing step of the nucleic acid workflow," (EX1005 (Gabriel Decl.) ¶ 248), all of McCloskey's examples use Sanger sequencing platforms such as an ABI 3100 with Big Dye. EX1012 (McCloskey) [0049], [0098]. McCloskey states that

“[m]ethods and conditions for analyzing nucleic acid molecules by gel electrophoresis, cloning, and sequencing are standard in the art” and “[e]xemplary conditions for analyzing bar-coded target nucleic acid molecule by gel electrophoresis, cloning, and sequencing are described in EXAMPLES 1-4 and 6.” EX1012 (McCloskey), at [0034]. A POSA would therefore understand that McCloskey’s methods were designed with Sanger sequencing in mind.

188. Industry literature cited by Dr. Gabriel supports the Petition’s characterization of Sanger sequencing as “labor intensive, [] only be partially automated, and [] slow and expensive.” Pet. at 4. Sanger sequencing was intended for small sample sizes and low numbers of target regions, because Sanger sequencing is “reliant on one-at-a-time PCR amplification of each target region.” EX1014 (Shendure) at 1141. Therefore, as outlined above, although Sanger sequencing can “achieve read-lengths of up to ~1,000 bp” long, Sanger sequencing is labor intensive and slow. EX1014 (Shendure) at 1135.
189. Bentley explains a single human genome “would take approximately 10,000 instrument days (e.g. 30 instruments for 1 year) to complete and would cost approximately \$10 million” using Sanger sequencing. A POSA would thus have understood that Sanger sequencing is costly and

often cost prohibitive for “most individual laboratories.” EX1014 (Shendure) at 1141 (“In a recent informal survey we conducted of the overall cost to conventionally sequence 100 genes from 100 samples, assuming each gene has an average of 10 exons, quoted estimates from noncommercial genome centers and commercial sequence service providers ranger from \$300,000 to \$1,000,000. Clearly, this cost is beyond the range of most individual laboratories.”).

190. Around the time of the invention, a POSA would have understood that “[c]ost and throughput are the key limitations of sequencing.” EX1017 (Bentley), at 545. Thus, at the time of the invention, a POSA reading McCloskey would thus be interested in keeping costs low and throughput high.

191. Conversely, Bignell is directed to next-generation sequencing. For example, Bignell uses next-generation sequencing platforms, such as the Illumina Genome Analyzer. EX1013 (Bignell), 41:51-61, 46:45-47:3, 45:48-80; EX1005 (Gabriel Decl.) ¶ 251. Next-generation sequencing technologies “reduced the cost of DNA sequencing by over two orders of magnitude, and democratiz[ed] the field by putting the sequencing capacity of a major genome center in the hands of individual investigators.” EX1014 (Shendure), at 1135, 1143. For example, Dr.

Gabriel explains that next-generation sequencing “made it possible to produce millions of sequencing reads in a single instrument run (compared to hundred with ‘first generation’ approaches).” EX1005 (Gabriel Decl.) ¶ 55. A POSA using next-generation sequencing would thus be looking for ways to limit costs and increase throughput.

192. Because large amounts of short reads are obtained in next-generation sequencing, the evolution of next-generation sequencing posed “challenges for bioinformatics” requiring the development of software specific for next-generation sequencing, such as “alignment tools ... developed specifically for rapid alignment of large sets of short reads.” EX1014 (Shendure), at 1142. A POSA would further understand that “the potential utility of short-read sequencing has been tremendously strengthened by the availability of whole genome assemblies for *Homo sapiens* and all major model organisms, as these effectively provide a reference against which short reads can be mapped.” EX1014 (Shendure), at 1135.

193. Dr. Gabriel states a major benefit of McCloskey is “adding unique barcodes to each nucleic acid molecule ... to identify and correct for errors and/or bias introduced by amplification.” EX1005 (Gabriel Decl.), ¶ 251 (citing EX1012 (McCloskey), [0005]). Dr. Gabriel explains that a

POSA would have recognized that “McCloskey’s unique barcodes with Bignell’s method of Next-Gen sequencing would have provided a solution to errors and/or bias caused by the amplification steps.” EX1005 (Gabriel Decl.) ¶ 251.

194. Dr. Gabriel states that “Next-Gen sequencing technologies were widely adopted because they provided sequencing methods which were both cost-effective and highly efficient” and “made sequencing more affordable for all types of experiments and analyses.” EX1005 (Gabriel Decl.) ¶ 251. Dr. Gabriel further asserts “Bignell’s method includes a number of amplification steps,” therefore “[a] person of ordinary skill in the art would have recognized that incorporating McCloskey’s unique barcodes with Bignell’s method of Next-Gen sequencing would have provided a solution to errors and/or bias caused by amplification steps.” EX1005 (Gabriel Decl.) ¶ 251. Dr. Gabriel’s analysis is wrong.

195. A POSA using next-generation sequencing technologies would have understood that “[c]ost and throughput are the key limitations of sequencing.” EX1017 (Bentley), 545. Dr. Gabriel asserts that a POSA would be motivated to use McCloskey’s unique barcodes but fails to appreciate the costs associated with unique barcodes. EX1005 (Gabriel Decl.) ¶¶ 247-257. The synthesis and incorporation of unique molecular

barcodes adds additional costs to a nucleic acid workflow. Those costs are easier to bear when dealing with the same sample sizes and low sequencing counts of McCloskey's Sanger sequencing. For example, adding a unique molecular barcode to each molecule in a sample of 150-200 sequences as was typical in Sanger sequencing (EX1012 (McCloskey), [0043]-[0104]¹) might only require 150-200 unique barcodes.

196. Conversely, adding a unique barcode to each of the millions of sequences in next-generation sequencing runs would result in much greater costs and expenses. A significant advantage of next generation sequencing was the ability to obtain millions more sequences in a single sequencing run than previous sequencing technologies. EX1005 (Gabriel Decl.) ¶55. Thus, a POSA would have understood that many more unique molecular barcodes would need to be synthesized when using next-generation sequencing platforms than when creating unique molecular barcodes for sequencing technologies like those used in McCloskey. EX1004 (Gabriel

¹ In example 1, 132 sequences were obtained from the experiment. In example 2, 46 sequences were obtained from the experiment. In example 3, 107 sequences were obtained from the experiment. In example 4, there number of sequencing reads was not reported. In example 5, there were no sequencing reads reported. In example 6, 142 barcodes were recovered.

Decl.) ¶ 55. Tellingly, none of McCloskey's examples recovered more than 150 barcodes in a single experiment, but Bignell suggests that millions of different sequences would need to be obtained to sequence an individual sample. EX1013 (Bignell), 3:28-38; EX1012 (McCloskey), [0043]-[0104].

197. Dr. Gabriel does not consider such costs at all, even though her cited references make clear that “[c]ost and throughput are the key limitations of sequencing.” EX1017 (Bentley), 545. Dr. Gabriel's failure to consider the prohibitive expense of applying McCloskey's method to Bignell's sequencing platform does not show motivation to combine the two references. Instead, a POSA would *not* have been motivated to incur the additional cost of synthesizing and incorporating random barcodes for millions of sequencing reads with next-generation sequencing technologies described in Bignell.

**XIII. GROUND 5: THE CHALLENGED CLAIMS ARE PATENTABLE
OVER MCLLOSKEY IN VIEW OF BIGNELL AND BENTLEY**

198. Dr. Gabriel asserts that Bignell, McCloskey, and Bentley render obvious claims 2 and 3. EX1005 (Gabriel Decl.) ¶¶ 307-327. Claims 2 and 3 depend on claim 1 but Dr. Gabriel does not provide any additional analysis related to claim 1 that includes Bentley. *See* EX1001, 131:49-

58; EX1005 (Gabriel Decl.) ¶¶ 307-327. For the reasons discussed above, the combination of McCloskey, Bignell and Bentley does not disclose claim 1. Section XII.A. For the same reasons I provide above in Section XII.B. for McCloskey and Bignell, a POSA would not have been motivated to combine McCloskey, Bignell, and Bentley.

**XIV. GROUND 6: THE CHALLENGED CLAIMS ARE PATENTABLE
OVER MCCLOSKEY IN VIEW OF BIGNELL AND GAHAN**

199. Dr. Gabriel asserts that Bignell, McCloskey, and Gahan render obvious claims 6 and 8. EX1005 (Gabriel Decl.) ¶¶ 328-334. Claims 6 and 8 depend on claim 1 but Dr. Gabriel does not provide any additional analysis related to claim 1 that includes Gahan. *See* EX1001, 132:30-37; EX1005 (Gabriel Decl.) ¶¶ 328-334 For the reasons discussed above, the combination of McCloskey, Bignell and Gahan does not disclose claim 1. Section XII.A. For the same reasons I provide above in Section XII.B for McCloskey and Bignell, a POSA would not have been motivated to combine McCloskey, Bignell, and Gahan.

**XV. GROUND 7: THE CHALLENGED CLAIMS ARE PATENTABLE
OVER MCCLOSEY IN VIEW OF BIGNELL AND CHIANG**

200. Dr. Gabriel asserts that Bignell, McCloskey, and Chiang renders obvious claim 7. EX1005 (Gabriel Decl.) ¶¶ 335-344. Claim 7 depends on claim 1 but Dr. Gabriel does not provide any additional analysis related to claim

1 that includes Chiang. *See* EX1001, 132:32-34; EX1005 (Gabriel Decl.) ¶¶ 335-344. For the reasons discussed above, the combination of McCloskey, Bignell and Chiang does not disclose claim 1. Section XII.A. For the same reasons I provide above in Section XII.B for McCloskey and Bignell, a POSA would not have been motivated to combine McCloskey, Bignell, and Chiang.

XVI. APPENDIX A: THE CHALLENGED CLAIMS OF THE '730 PATENT

1. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

- a. obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
- b. introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;
- c. amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
- d. sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
- e. correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

2. The method of claim 1, wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences.

3. The method of claim 2, further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences.

4. The method of claim 1, wherein said biological sample is selected from the group consisting of blood and biopsy tissue.

5. The method of claim 1, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.

6. The method of claim 1, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.

7. The method of claim 1, further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained.

8. The method of claim 1, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

9. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:

- a. obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
- b. introducing a set of unique differentiator tags, wherein members of said set of unique differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin;
- c. amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
- d. sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
- e. correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule.

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true, and further that these 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.

July 18, 2025

Date

Joseph D. Puglisi
Joseph D. Puglisi, Ph.D.

Joseph D. Puglisi, PhD

Jauch Professor and Professor of Structural Biology

Stanford University School of Medicine

Stanford, CA 94305-5126

Tel. (650) 498-4397 or (650) 723-9151

FAX (650) 723-8464

E-mail: puglisi@stanford.edu

*****puglisi.stanford.edu

Academic History:**Colleges and Universities attended:**

1984-1989 University of California, Berkeley, Ph.D. in Biophysical Chemistry.

1980-1984 The Johns Hopkins University, B.A. in Chemistry.

Scholarships and Honors:

1997 Alfred P. Sloan Research Fellow.

1994 David and Lucille Packard Fellowship in Science and Engineering.

1993 Camille and Henry Dreyfus Teacher-Scholar Award.

1989 EMBO Fellow.

1989 Thesis nominated for J.T. Baker Award in Chemistry.

1989 Sigma Xi Honor Society.

1988 Department of Chemistry, Berkeley. Bruce Mahan Memorial Teaching Award.

1987 Outstanding Graduate Teaching Assistant.

1984-1985 Department of Chemistry, Berkeley. Bruce Mahan Memorial Teaching Award.

2011 NIH Merit Award.

2011 NIH Director's Transformative R01 (T-R01) Program Award, NIH.

2014 Member, National Academy of Sciences.

Named Lectures

2014 Linus Pauling Lecture, Stanford University

2018 David L. Weaver Lecture, UC Davis

2021 Tom Steitz Memorial Lecture, Yale University

2021 RiboClub Canada, Keynote Lecture

2022 Ferry Lectureship in Polymer Science, University of Wisconsin

2022 Michael O'Keefe Memorial Lecture, Arizona State University

2022 Ribosome Meeting. Keynote Lecture

Postdoctoral Training:

1991-1993 Postdoctoral Fellow, Department of Chemistry
Massachusetts Institute of Technology, Cambridge, MA
(Advisor: James R. Williamson).

1989-1991 Postdoctoral Fellow; EMBO Postdoctoral Fellowship
I.B.M.C du CNRS, Strasbourg, France. (Advisor: Richard Giegé).

Employment History:

9/2004-Present Professor of Structural Biology

2024-present Vice Chair, Department of Structural Biology

9/2004-9/2014	Chair, Department of Structural Biology
4/2004-9/2004	Acting Chair, Department of Structural Biology
1998-3/2004	Associate Chair, Department of Structural Biology, Stanford Univ. School of Medicine.
7/1/97-Present	Director, Stanford Magnetic Resonance Laboratory (SMRL).
7/97-8/2004	Associate Professor of Structural Biology, Stanford Univ. School of Medicine.
1993-1997	Affiliate Faculty, Molecular, Cellular, & Developmental Biology Graduate Program in Biology, University of California, Santa Cruz.
1997	Associate Professor of Chemistry and Biochemistry University of California, Santa Cruz.
1993-1997	Assistant Professor of Chemistry and Biochemistry University of California, Santa Cruz.
1984-1989	Research Assistant, University of California, Berkeley Advisor: Professor Ignacio Tinoco, Jr.
1983-1984	Undergraduate Research Fellow, Organic Synthesis, The Johns Hopkins University, Advisor: Professor Alex Nickon.
1981-1984	Summer Research Scientist, Airco Central Research Laboratory, Murray Hill, N.J.

Research Interests:

The Puglisi group investigates the role of RNA in health and disease. Our goal is to understand the dynamic interplay of RNA conformation, function and mechanism using tools of molecular biophysics, structural biology, physical chemistry and biochemistry. We apply structural methods such as NMR) spectroscopy, x-ray crystallography and cryoEM and integrate structural views with real time dynamics at the single-molecule level. Our goal is a unified picture of structure, dynamics and function. A long-term goal is to target processes involving RNA with novel therapeutic strategies.

Jody Puglisi is a leading scientist in the study of RNA structure and function in biology. He has pioneered the use of NMR in the study of RNAs and RNA-ligand complexes. His early work showed how small molecules can bind to RNAs, which has lead to the field of RNA as a drug target and riboswitches. Recent work has merged novel biophysical, biochemical and structural methods to merge dynamic and static structural views of essential RNA mechanisms. He has a deep interest in nucleic acid sequencing and analysis technologies over the past two decades.

Current research areas include:

- Fundamental mechanisms of translation initiation, elongation, termination and recycling
- Ribosomal pausing induced by RNA modifications, RNA structure and nascent peptides and proteins
- Recoding events, including frameshifting and bypassing
- Drug-ribosome interactions and mechanism
- Mechanism of eukaryotic translation initiation—5' end recognition, scanning, start codon recognition
- HCV IRES mediated translation initiation
- 3' end regulation of translation
- Role of translational processes in human disease
- Single-molecule methods

Professional Service:

07/2024 Organizer, ASI 20th Course, ISBMR, Erice Sicily, Italy “Frontiers of Biophysics”

07/2023	Organizer, ASI 19 TH Course, ISBMR, Erice, Sicily, Italy: "Frontiers of Biophysics." ***** puglisilab.org/erice2023.html
07/2022	Organizer, ASI 18 TH Course, ISBMR, Erice, Sicily, Italy: "Frontiers of Biophysics." ***** puglisilab.org/erice2022.html
08/2019	Organizer, ASI 17 TH Course, ISBMR, Erice, Sicily, Italy: "Frontiers of Biophysics." ***** puglisilab.org/erice2019.html
08/2018	Organizer, ASI 16 TH Course, ISBMR, Erice, Sicily, Italy: "Frontiers of Biophysics." ***** puglisilab.org/erice2018.html
05/2017	Organizer, ASI 15 TH Course, ISBMR, Erice, Sicily, Italy: "Biophysics and Molecular Structure." ***** smrl.stanford.edu/erice2017/
05/2016	Organizer, ASI 14 TH Course, ISBMR, Erice, Sicily, Italy: "Future of Molecular Biophysics." ***** smrl.stanford.edu/erice2016/
07/2014	Organizer, ASI 13 TH Course, ISBMR, Erice, Sicily, Italy: "Future of Biophysics and Structural Biology." ***** smrl.stanford.edu/erice2014/
02/2014	Council Member, Biophysical Society.
06/2013	Organizer, ASI 12 TH Course, ISBMR, Erice, Sicily, Italy: "Future of Biophysics." ***** smrl.stanford.edu/erice2013/
02/2013	Program Chair, 57 th Annual Biophysical Society National Meeting.
06/2012	Organizer, ASI 11 TH Course, ISBMR, Erice, Sicily, Italy: "Frontiers of Biophysics and Structural Biology." http://smrl.stanford.edu/erice2012/
2011-Present	Member, Advisory Editorial Board, EMBO reports.
06/2010	Organizer, NATO-ASI 10 TH Course, ISBMR, Erice, Sicily, Italy: "Biophysics and Structure to Counter Threats and Challenges." ***** smrl.stanford.edu/erice2010/
06/2009	Organizer, ASI 9 TH Course, International School of Biological Magnetic Resonance, Erice, Sicily, Italy: "Biophysics and Structure to Counter Threats and Challenges." ***** smrl.stanford.edu/erice2009/
06/2007	Organizer, NATO-ASI 8 TH Course, ISBMR, Erice, Sicily, Italy: "Biophysics and the Challenges of Emerging Threats." ***** smrl.stanford.edu/erice2007/
06/2007	Co-Chair, Nucleic Acids Gordon Conference, Salve Regina University, Newport, RI.
03/2007	Co-Chair, ACS National Meeting, RNA Biophysics Symposium, Chicago, IL
06/2006	Co-Chair, Nucleic Acids Meeting, Salve Regina University, Newport, RI.
2005-2009	Chair, Provost's Advisory Board for Postdoctoral Scholars, Stanford Univ.
06/2005	Organizer, NATO-ASI 7 TH Course, ISBMR, Erice, Sicily, Italy: "Structure and Biophysics - New Technologies for Current Challenges in Biology and Beyond." ***** smrl.stanford.edu/erice2005/
10/2003-	Senior Editor, Structure (Cell Press).
07/2003-	Director, International School of Biological Magnetic Resonance (ISBMR), Ettore Majorana Centre for Scientific Culture, Erice, Italy.
07/2003	Organizer, NATO-ASI 6 TH Course: Structure, Dynamics and Function of Biological Macromolecules and Assemblies. ISBMR, Erice, Sicily, Italy. ***** smrl.stanford.edu/erice2003/
05/2003-	Member, Fellowships Award Committee, Stanford Univ. School of Medicine.
2003-2007	NIH Molecular Biophysics and Cellular Biophysics Study Section, Center for Scientific Review (BBCA).
2002-2005	Member, Provost's Advisory Board for Postdoctoral Scholars, Stanford Univ.
2002-2005	Chair, Postdoctoral Affairs Committee, Stanford Univ. School of Medicine.
2001-2002	Member, Postdoctoral Affairs Committee, Stanford Univ. School of Medicine.
06/2001	Co-organizer, NATO ASI, 5 th Course: Protein Structure, Dynamics, Genomics and Function. Int'l School of Biological Magnetic Resonance (ISBMR), Erice, Sicily, Italy. ***** smrl.stanford.edu/erice2001/

10/2000 BBICA Study Session Ad Hoc.
10/1999 NIH Review Panel, Physiological Chemistry Study Session Ad Hoc.
10/1998- Program in Molecular & Genetic Medicine Advisory Board (PMGM)
 Stanford Univ. School of Medicine.
10/1989- Journal Reviewer. Regularly reviews for Science, Nature, Nature Structural Biology, Cell,
 Molecular Cell, J. Mol. Biol., J. Mag Res., J. Biol. NMR, Biochemistry, J. Am. Chem.
 Soc., Nucleic Acids Research.

Peer-Reviewed Articles:

186. Grosely R., Alvarado, C, Ivalyo IP, Nicholson OB, **Puglisi JD**, Dever TD Lapointe CD (2025). eIF1 and eIF5 dynamically control translation start site fidelity. *Nature Struct. Mol. Biol* (in press)
185. **Puglisi JD** (2025). Taming the Ribosome. *Biophys J*. 2024 Sep 17;123(18):2964-2965. doi: 10.1016/j.bpj.2024.07.045. Epub 2024 Aug 2. PMID: 39097772
184. Krahn N, Zhang J, Melnikov SV, Tharp JM, Villa A, Patel A, Howard RJ, Gabir H, Patel TR, Stetefeld J, **Puglisi J**, Söll D (2024). tRNA shape is an identity element for an archaeal pyrrolysyl-tRNA synthetase from the human gut. *Nucleic Acids Res.*52(2):513-524. doi: 10.1093/nar/gkad1188. PMID: 3810036. PMCID: 10810272
183. Guca E, Alarcon R, Palo MZ, Santos L, Alonso-Gil S, Davyt M, de Lima LHF, Boissier F, Das S, Zagrovic B, **Puglisi JD**, Hashem Y, Ignatova Z. (2024). N⁶-methyladenosine in 5'UTR does not promote translation initiation. *Mol. Cell* Feb 1;84(3):584-595.e6. doi: 10.1016/j.molcel.2023.12.028. Epub 2024 Jan 19. PMID: 38244546. PMCID: 10576065
182. Huang T, Choi J, Prabhakar A, **Puglisi JD**, Petrov A. (2023). Partial spontaneous intersubunit rotations in pretranslocation ribosomes. *Proc Natl Acad Sci U S A* 2023 Oct 10;120(41):e2114979120. doi: 10.1073/pnas.2114979120. Epub 2023 Oct 6. PMID 37801472. PMCID: 10576065
181. Nouhin J, Tzou PL, Rhee SY, Sahoo MK, Pinsky BA, Krupkin M, **Puglisi JD**, Puglisi EV, Shafer RW. (2023). Human immunodeficiency virus 1 5'-leader mutations in plasma viruses before and after the development of reverse transcriptase inhibitor-resistance mutations. *J. Gen. Virol.* 2023 Oct;104(10).doi: 10.1099/jgv.0.001898. PMID 37801004. PMCID: 10721937
180. Schubert K, Karousis ED, Ban I, Lapointe CP, Leibundgut M, Bäuml E, Kummerant E, Scaiola A, Schönhut T, Ziegelmeüller J, **Puglisi JD**, Mühlemann O, Ban N. (2023). Universal features of Nsp1-mediated translational shutdown by coronaviruses. *Mol Cell* Oct 5;83(19):3546-3557.e8. doi: 10.1016/j.molcel.2023.09.002. PMID: 37802027. PMCID:10312502
179. Wang J, Koduru T, Harish B, McCallum SA, Larsen KP, Patel KS, Peters EV, Gillilan RE, Puglisi EV, **Puglisi JD**, Makhatadze G, Royer CA. (2023). Pressure pushes tRNA^{Lys3} into excited conformational states. *Proc Natl Acad Sci U S A* 2023 Jun 27;120(26):e2215556120. doi: 10.1073/pnas.2215556120. Epub 2023 Jun 20. PMCID: 10293818
178. Hentschel J, Badstübner M, Choi J, Bagshaw CR, Lapointe CP, Wang J, Jansson LI, **Puglisi JD**, Stone MD. (2023). Real-time detection of human telomerase DNA synthesis by multiplexed single-molecule FRET. *Biophys. J.* 2023 Jul 27:S0006-3495(23)00471-X. doi: 10.1016/j.bpj.2023.07.019 Online ahead of print. PMID: 37515327. PMCID: 10502476
177. Prabhakar A, Pavlov MY, Zhang J, Indrisiunaite G, Wang J, Lawson MR, Ehrenberg M, **Puglisi JD**. (2023). Dynamics of release factor recycling during translation termination in bacteria. *Nucleic Acids Res.* 2023 Apr 27;gkad286. doi: 10.1093/nar/gkad286. Online ahead of print. PMID: 37102635. PMCID: PMC10287982
176. Wang J, Shin BS, Alvarado C, Kim JR, Bohlen J, Dever TE, **Puglisi JD**. (2022). Rapid 40S scanning and its regulation by mRNA structure during eukaryotic translation initiation. *Cell*. Nov 23;185(24):4474-4487.e17. doi: 10.1016/j.cell.2022.10.005. Epub 2022 Nov 4. PMID: 36334590
175. Prabhakar A, Krahn N, Zhang J, Vargas-Rodriguez O, Krupkin M, Fu Z, Acosta-Reyes FJ, Ge X, Choi J, Crnković A, Ehrenberg M, Puglisi EV, Söll D, **Puglisi J**. (2022). Uncovering translation

- roadblocks during the development of a synthetic tRNA. *Nucleic Acids Res.* Oct 14;50(18):10201-10211. doi: 10.1093/nar/gkac576.PMID: 35882385
174. Lapointe CP, Grosely R, Sokabe M, Alvarado C, Wang J, Montabana E, Villa N, Shin BS, Dever TE, Fraser CS, Fernández IS, **Puglisi JD** (2022). eIF5B and eIF1A reorient initiator tRNA to allow ribosomal subunit joining. *Nature*. Jul;607(7917):185-190. doi: 10.1038/s41586-022-04858-z. Epub 2022 Jun 22.PMID: 35732735
 173. Lawson MR, Lessen LN, Wang J, Prabhakar A, Corsepius NC, Green R, **Puglisi JD**.(2021).Mechanisms that ensure speed and fidelity in eukaryotic translation termination. *Science* Aug 20;373(6557):876-882. doi: 10.1126/science.abi7801.PMID: 34413231.
 172. Leong K, Indrisiunaite G, Prabhakar A, **Puglisi JD**, Ehrenberg M. (2021). N 6-Methyladenosines in mRNAs reduce the accuracy of codon reading by transfer RNAs and peptide release factors. *Nucleic Acids Res.* Mar 18;49(5):2684-2699. doi: 10.1093/nar/gkab033.
 171. Lapointe CP, Grosely R, Johnson AG, Wang J, Fernández IS, **Puglisi JD**. (2021). Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation. *Proc Natl Acad Sci U S A.* Feb 9;118(6):e2017715118. Doi 10.1073/pnas.2017715118.PMID: 33479166. PMC8017934
 170. Wang J, Wang J, Shin BS, Kim JR, Dever TE, **Puglisi JD**, & Fernandez IS (2020). Structural basis for the transition from initiation to elongation by an 80S-eIF5B complex. *Nature Communications* Oct 6;11(1):5003. doi: 10.1038/s41467-020-18829-3.
 169. Johnson, AG, Flynn, RA, Lapointe, CP, Ooi, YS, Zhao, ML, Richards, CM, Qiao, W, Yamada, SB, Couthouis, J, Gitler, AD, Carette, JE, & **Puglisi, JD**. (2020). A memory of eS25 loss drives resistance phenotypes. *Nucleic Acids Research*. Jul 27;48(13):7279-7297. doi: 10.1093/nar/gkaa444.PMID: 32463448; PMC7367175
 168. O'Loughlin, S, Capece, MC, Klimova, M, Wills, NM, Coakley, A, Samatova, E, O'Connor, P, Loughran, G, Weissman, JS, Baranov, PV, Rodnina, MV, **Puglisi, JD**, & Atkins, JF. (2020). Polysomes Bypass a 50-Nucleotide Coding Gap Less Efficiently Than Monosomes Due to Attenuation of a 5' mRNA Stem-Loop and Enhanced Drop-off. *Journal of molecular biology*, **S0022-2836(20)**30346-6. Advance online publication. doi.org/10.1016/j.jmb.2020.05.010. PMC7245289.
 167. Choi J, O'Loughlin S, Atkins JF, **Puglisi JD**. (2020). The energy landscape of -1 ribosomal frameshifting. *Sci Adv.* 2020 Jan 1;**6(1)**:eaax6969. doi: 10.1126/sciadv.aax6969. eCollection 2020 Jan. PMCID: PMC6938710.
 166. Chandrasekaran, V, Juszkievicz, S, Choi, J, **Puglisi, JD**, Brown, A, Shao, S, Ramakrishnan, V, & Hegde, RS. (2019). Mechanism of ribosome stalling during translation of a poly(A) tail. *Nature structural & molecular biology*, **26(12)**, 1132–1140. doi.org/10.1038/s41594-019-0331-x. PMC6900289.
 165. Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, Szczesny P, Piston DW, Zaher HS, **Puglisi JD**, Djuranovic S. (2019). A short translational ramp determines the efficiency of protein synthesis. *Nat Commun.* 2019 Dec 18;**10(1)**:5774. doi: 10.1038/s41467-019-13810-1. PMCID: PMC6920384.
 164. Choi J, Marks J, Zhang J, Chen DH, Wang J, Vázquez-Laslop N, Mankin AS, **Puglisi JD**. (2019). Dynamics of the context-specific translation arrest by chloramphenicol and linezolid. *Nat Chem Biol.* 2019 Dec 16. doi: 10.1038/s41589-019-0423-2. [Epub ahead of print] PMID: 31844301.

163. Duss O, Stepanyuk GA, **Puglisi JD**, Williamson JR. (2019). Transient Protein-RNA Interactions Guide Nascent Ribosomal RNA Folding. *Cell*. 2019 Nov 27;**179(6)**:1357-1369.e16. doi: 10.1016/j.cell.2019.10.035. Epub 2019 Nov 21. PMID: 31761533.
162. Wang J, Johnson AG, Lapointe CP, Choi J, Prabhakar A, Chen DH, Petrov AN, **Puglisi JD**. (2019). eIF5B gates the transition from translation initiation to elongation. *Nature*. 2019 Sep;**573(7775)**:605-608. doi: 10.1038/s41586-019-1561-0. Epub 2019 Sep 18. PMCID: PMC6763361.
161. Yamada SB, Gendron TF, Niccoli T, Genuth NR, Grosely R, Shi Y, Glaria I, Kramer NJ, Nakayama L, Fang S, Dinger TJI, Thoeng A, Rocha G, Barna M, **Puglisi JD**, Partridge L, Ichida JK, Isaacs AM, Petrucelli L, Gitler AD. (2019). RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci*. 2019 Sep;**22(9)**:1383-1388. doi: 10.1038/s41593-019-0455-7. Epub 2019 Jul 29. PMC6713615.
160. Larsen KP, Choi J, Prabhakar A, Puglisi EV, **Puglisi JD**. (2019). Relating Structure and Dynamics in RNA Biology. *Cold Spring Harb Perspect Biol*. 2019 Jul 1;**11(7)**. pii: a032474. doi: 10.1101/cshperspect.a032474. Review. PMID:31262948.
159. Choi J, Grosely R, Puglisi EV, **Puglisi JD**. (2019). Expanding single-molecule fluorescence spectroscopy to capture complexity in biology. *Curr Opin Struct Biol*. 2019 Jun 15. pii: S0959-440X(18)30129-5. doi: 10.1016/j.sbi.2019.05.005. [Epub ahead of print] Review. PMID:31213390.
158. Johnson AG, Lapointe CP, Wang J, Corsepius NC, Choi J, Fuchs G, **Puglisi JD**. (2019). RACK1 on and off the ribosome. *RNA*. 2019 Jul;**25(7)**:881-895. doi: 10.1261/rna.071217.119. Epub 2019 Apr 25. PMC6573788.
157. Prabhakar A, Puglisi EV, **Puglisi JD**. (2019). Single-Molecule Fluorescence Applied to Translation. *Cold Spring Harb Perspect Biol*. 2019 Jan 2;**11(1)**. pii: a032714. doi: 10.1101/cshperspect.a032714. Review. PMC6314065.
156. Duss O, Stepanyuk GA, Grot A, O'Leary SE, **Puglisi JD**, Williamson JR. (2018). Real-time assembly of ribonucleoprotein complexes on nascent RNA transcripts. *Nat Commun*. 2018 Nov 30;**9(1)**:5087. doi: 10.1038/s41467-018-07423-3. PMC6269517.
155. Coey AT, Larsen KP, Choi J, Barrero DJ, **Puglisi JD**, Puglisi EV. (2018). Dynamic Interplay of RNA and Protein in the Human Immunodeficiency Virus-1 Reverse Transcription Initiation Complex. *J Mol Biol*. 2018 Dec 7;**430(24)**:5137-5150. doi: 10.1016/j.jmb.2018.08.029. Epub 2018 Sep 7. PMC6289658.
154. Kappel K, Liu S, Larsen KP, Skiniotis G, Puglisi EV, **Puglisi JD**, Zhou ZH, Zhao R, Das R. (2018). De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat Methods*. 2018 Nov;**15(11)**:947-954. doi: 10.1038/s41592-018-0172-2. Epub 2018 Oct 30. PMC6636682.
153. Choi J, Grosely R, Prabhakar A, Lapointe CP, Wang J, **Puglisi JD**. (2018). How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. *Annu Rev Biochem*. 2018 Jun 20;**87**:421-449. doi: 10.1146/annurev-biochem-060815-014818. Review. PMC6594189.
152. Larsen KP, Mathiharan YK, Kappel K, Coey AT, Chen DH, Barrero D, Madigan L, **Puglisi JD**, Skiniotis G, Puglisi EV. (2018). Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature*. 2018 May;**557(7703)**:118-122. doi: 10.1038/s41586-018-0055-9. Epub 2018 Apr 25. PMC5934294.
151. Choi J, Indrisiunaite G, DeMirci H, Jeong KW, Wang J, Petrov A, Prabhakar A, Rechavi G, Dominissini D, He C, Ehrenberg M, **Puglisi JD**. (2018). 2'-O-methylation in mRNA disrupts tRNA

- decoding during translation elongation. *Nat Struct Mol Biol.* 2018 Mar;25(3):208-216. doi: 10.1038/s41594-018-0030-z. Epub 2018 Feb 19. PMID: PMC5840002
150. Johnson AG, Petrov AN, Fuchs G, Majzoub K, Grosely R, Choi J, **Puglisi JD.** (2018). Fluorescently-tagged human eIF3 for single molecule spectroscopy. *Nucleic Acids Res.* 2018 Jan 25;46(2):e8. doi: 10.1093/nar/gkx1050. PMID: PMC5778468.
149. Choi J, **Puglisi JD.** (2017). Three tRNAs on the ribosome slow translation elongation. *Proc Natl Acad Sci U S A.* Dec 26;114(52):13691-13696. doi: 10.1073/pnas.1719592115. Epub 2017 Dec 11. PMID: PMC5748233.
148. Prabhakar, A. Capece, MC, Petrov, A, Choi, J & **Puglisi JD.** (2017). Post-termination ribosome intermediate acts as the gateway to ribosome recycling. *Cell Reports.* Jul 5;20(1):161-172. doi: 10.1016. PMID: 28683310. PMID: PMC5555083.
147. Prabhakar A, Choi J, Wang J, Petrov A, **Puglisi JD.** (2017). Dynamic basis of fidelity and speed in translation: Coordinated multistep mechanisms of elongation and termination. *Protein Sci.* 2017 May 8. doi: 10.1002/pro.3190. [Epub ahead of print] Review. PMID: 28480640. PMID: PMC5477533.
146. Johnson AG, Grosely R, Petrov AN, **Puglisi, JD** (2017). Dynamics of IRES-mediated translation. *Philos Trans R Soc Lond B Biol Sci.* 2017 Mar 19;372(1716). pii: 20160177. doi: 10.1098/rstb.2016.0177. Review. PMID: 28138065. PMID: PMC5311923.
145. Chen J, Choi J, O'Leary SE, Prabhakar A, Petrov A, Grosely R, Puglisi EV, **Puglisi JD.** (2016). The molecular choreography of protein synthesis: translational control, regulation, and pathways. *Q Rev Biophys.* 2016 Jan;49:e11. PMID: 27658712.
144. Coey A, Larsen K, **Puglisi JD,** Puglisi EV. (2016) Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *RNA.* 2016 Sep 9. [Epub ahead of print] PMID: 27613581. PMID: PMC5066621.
143. Navon SP, Kornberg G, Chen J, Schwartzman T, Tsai A, Puglisi EV, **Puglisi JD,** Adir N. (2016). Amino acid sequence repertoire of the bacterial proteome and the occurrence of untranslatable sequences. *Proc Natl Acad Sci U S A.* 2016 Jun 28;113(26):7166-70. doi: 10.1073/pnas.1606518113. PMID: 27307442. PMID: PMC4932979.
142. Petrov A, Grosely R, Chen J, O'Leary SE, **Puglisi JD.** (2016). Multiple Parallel Pathways of Translation Initiation on the CrPV IRES. *Mol Cell.* 2016 Apr 7;62(1):92-103. doi: 10.1016/j.molcel.2016.03.020. PMID: 27058789. PMID: PMC4826567.
141. Tsai A, **Puglisi JD,** Uemura S. (2016). Probing the Translation Dynamics of Ribosomes Using Zero-Mode Waveguides. *Prog Mol Biol Transl Sci.* 2016;139:1-43. doi: 10.1016/bs.pmbts.2015.10.006. PMID: 26970189.
140. Choi J, Jeong KW, Demirci H, Chen J, Petrov A, Prabhakar A, O'Leary SE, Dominissini D, Rechavi G, Soltis SM, Ehrenberg M, **Puglisi JD.** (2016). N(6)-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat Struct Mol Biol.* **23(2)**:110-5. doi: 10.1038/nsmb.3148. Epub 2016 Jan 11. PMID: 26751643. PMID: PMC4826618.
139. Sierra RG, Gati C, Laksmono H, Dao EH, Gul S, Fuller F, Kern J, Chatterjee R, Ibrahim M, Brewster AS, Young ID, Michels-Clark T, Aquila A, Liang M, Hunter MS, Koglin JE, Boutet S, Junco EA, Hayes B, Bogan MJ, Hampton CY, Puglisi EV, Sauter NK, Stan CA, Zouni A, Yano J, Yachandra VK, Soltis SM, **Puglisi JD,** DeMirici H. (2016) Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nat Methods.* **13(1)**:59-62. doi: 10.1038/nmeth.3667. Epub 2015 Nov 30. PMID: 26619013. PMID: PMC4890631.

138. Siprashvili Z, Webster DE, Johnston D, Shenoy RM, Ungewickell AJ, Bhaduri A, Flockhart R, Zarnegar BJ, Che Y, Meschi F, **Puglisi JD**, Khavari PA. (2016). The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. *Nat Genet.* **48(1)**:53-8. doi: 10.1038/ng.3452. Epub 2015 Nov 23. PMID: 26595770. PMC5324971.
137. Chen J, Coakley A, O'Connor M, Petrov A, O'Leary SE, Atkins JF, **Puglisi JD**. (2015) Coupling of mRNA Structure Rearrangement to Ribosome Movement during Bypassing of Non-coding Regions. *Cell.* **163(5)**:1267-80. doi: 10.1016/j.cell.2015.10.064. PMID: 26590426.
136. Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, Johansson M, Müller-Lucks A, Trovato F, **Puglisi JD**, O'Brien EP, Beckmann R, von Heijne G. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **12(10)**:1533-40. doi: 10.1016/j.celrep.2015.07.065. Epub 2015 Aug 28. PMID: 26321634. PMCID: PMC4571824.
135. **Puglisi, JD.** (2015) Synthetic biology: Ribosomal ties that bind. *Nature.* **524(7563)**:45-6. doi: 10.1038/nature14642. Epub 2015 Jul 29. PMID: 26222027.
134. **Puglisi JD.** (2015) Protein synthesis. The delicate dance of translation and folding. *Science.* **348(6233)**:399-400. doi: 10.1126/science.aab2157. PMID: 25908811.
133. **Puglisi JD.** (2015) RNA dances to center stage. *RNA.* **21(4)**:712-3. doi: 10.1261/rna.051078.115. PMC4371346.
132. Capece MC, Kornberg GL, Petrov A, **Puglisi JD.** (2015) A simple real-time assay for in vitro translation. *RNA.* **21(2)**:296-305. PMID: 25525154; PMC4338355
131. Fuchs, G., Petrov, A. N., Marceau, C. D., Popov, L. M., Chen, J., O'Leary, S. E., Wang, R., Carette, J. E., Sarnow, P., **Puglisi, JD.** (2015) Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proceedings of the National Academy of Sciences of the United States of America.* **112 (2)**:319-325. PMID: 25516984
130. Noriega, T. R., Chen, J., Walter, P., **Puglisi, JD.** (2014) Real-time observation of signal recognition particle binding to actively translating ribosomes *ELIFE*, Oct 30;3. PMID: 25358118; PMC4213662.
129. Chen J, Petrov A, Johansson M, Tsai A, O'Leary SE, **Puglisi JD.** (2014). Dynamic pathways of -1 translational frameshifting. *Nature.* 2014 Jun 11. doi: 10.1038/nature13428. PMID: 24919156; PMC4472451.
128. Tsai A, Kornberg G, Johansson M, Chen J, **Puglisi JD.** (2014). The Dynamics of SecM-Induced Translational Stalling. *Cell Rep.* **7(5)**:1521-33. PMC4059775.
127. Johansson M, Chen J, Tsai A, Kornberg G, **Puglisi JD.** (2014). Sequence-dependent elongation dynamics on macrolide-bound ribosomes. *Cell Rep.* **7(5)**:1534-46. PMID: 24836000; PMC4387896
126. Noriega TR, Tsai A, Elvekrog MM, Petrov A, Neher SB, Chen J, Bradshaw N, **Puglisi JD**, Walter P. (2014). Signal Recognition Particle-Ribosome Binding Is Sensitive To Nascent-Chain Length. *J Biol Chem.* pii: jbc.M114.563239. [Epub ahead of print] PMID: 24808175; PMC4094042
125. Chen J, Dalal RV, Petrov AN, Tsai A, O'Leary SE, Chapin K, Cheng J, Ewan M, Hsiung PL, Lundquist P, Turner SW, Hsu DR, **Puglisi JD.** (2014). High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proc Natl Acad Sci U S A.* **111(2)**:664-9. PMID: 24379388; PMC4387896
124. O'Leary SE, Petrov A, Chen J, **Puglisi JD.** (2013). Dynamic recognition of the mRNA cap by *Saccharomyces cerevisiae* eIF4E. *Structure.* **21(12)**:2197-207. PMCID: PMC3878992.
123. Petrov A, Wu T, Puglisi EV, **Puglisi JD.** (2013). RNA purification by preparative polyacrylamide gel electrophoresis. *Methods Enzymol.* **530**:315-30. PMID: 24034329.

122. Petrov A, Tsa A, **Puglisi JD.** (2013). Analysis of RNA by analytical polyacrylamide gel electrophoresis. *Methods Enzymol.* **530**:301-13. PMID: 24034328.
121. Simonetti A, Marzi S, Billas IM, Tsai A, Fabbretti A, Myasnikov AG, Roblin P, Vaiana AC, Hazemann I, Eiler D, Steitz TA, **Puglisi JD,** Gualerzi CO, Klaholz BP. (2013). Involvement of protein IF2 N domain in ribosomal subunit joining revealed from architecture and function of the full-length initiation factor. *Proc Natl Acad Sci U S A.* **110(39)**:15656-61. PMCID: PMC3785770.
120. Chen J, Petrov A, Tsai A, O'Leary SE, **Puglisi JD.** (2013) Coordinated conformational and compositional dynamics drive ribosome translocation. *Nat Struct Mol Biol*; **20(6)**:718-27. PMCID: PMC3883222
119. Tsai A, Uemura S, Johansson M, Puglisi EV, Marshall RA, Aitken CE, Korlach J, Ehrenberg M, **Puglisi JD.** (2013). The Impact of Aminoglycosides on the Dynamics of Translation Elongation. *Cell Rep.* **3(2)**:497-508. PMID: 23416053.
118. Chen J, Tsai A, O'Leary SE, Petrov A, **Puglisi JD.** (2012). Unraveling the dynamics of ribosome translocation. *Curr Opin Struct Biol.* **22 (6)**:804-14. PMID: 23142574.
117. Tsai A, Petrov A, Marshall RA, Korlach J, Uemura S, **Puglisi JD.** (2012). Heterogeneous pathways and timing of factor departure during translation initiation. *Nature.* **487(7407)**:390-3. PMID: 22722848.
116. Harvey CJ, **Puglisi JD,** Pande VS, Cane DE, Khosla C. (2012). Precursor directed biosynthesis of an orthogonally functional erythromycin analogue: selectivity in the ribosome macrolide binding pocket. *J Am Chem Soc.* **134(29)**:12259-65. PMID: 22741553.
115. Petrov A, Chen J, O'Leary S, Tsai A, **Puglisi JD.** (2012). Single-Molecule Analysis of Translational Dynamics. *Cold Spring Harb. Perspect Biol.* 4:a011551. PMID: 22798542
114. **Puglisi JD,** Williamson JR. (2012). Digging deep into nucleic acid structure and nucleic acid protein recognition. *Curr Opin Struct Biol.* **22(3)**:249-50. PMID: 22632873.
- ☐ 113. Chen J, Tsai A, Petrov A, **Puglisi JD.** (2012). Nonfluorescent Quenchers To Correlate Single-Molecule Conformational and Compositional Dynamics. *J Am Chem Soc.* **134(13)**:5734-7. PMID: 22428667.
- ☐ 112. Masuda T, Petrov AN, Iizuka R, Funatsu T, **Puglisi JD,** Uemura S. (2012). Initiation factor 2, tRNA, and 50S subunits cooperatively stabilize mRNAs on the ribosome during initiation. *Proc Natl Acad Sci USA.* **109(13)**:4881-5. PMID: 22411833.
111. Puglisi EV, **Puglisi JD.** (2011). Secondary structure of the HIV reverse transcription initiation complex by NMR. *J Mol Biol.* **410(5)**:863-74. PMID: 21763492.
110. Petrov A, Kornberg G, O'Leary S, Tsai A, Uemura S, **Puglisi JD.** (2011). Dynamics of the translational machinery. *Curr Opin Struct Biol.* **21(1)**:137-45. Review. PMID: 21256733.
109. Aitken CE, **Puglisi JD.** (2010). Following the intersubunit conformation of the ribosome during translation in real time. *Nat Struct Mol Biol.* **17(7)**:793-800. PMID: 20562856.
108. **Puglisi JD,** Williamson JR (2010). Nucleic acids continue to surprise. *Curr Opin Struct Biol.* **20(3)**:259-61. PMID: 20510604.
107. Petrov A, **Puglisi JD.** (2010) Site-specific labeling of *Saccharomyces cerevisiae* ribosomes for single-molecule manipulations. *Nucleic Acids Res.* **38(13)**:e143. PMID: 20501598.
106. Uemura, S., Aitken, CE, Korlach, J., Flusberg, BA, Turner, SW & **Puglisi, JD.** (2010). Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature.* **464(7291)**:1012-7. PMID: 20393556.

105. Aitken CE, Petrov A, **Puglisi JD.** (2010). Single Ribosome Dynamics and the Mechanism of Translation. *Annu Rev Biophys.* **39**:491-513. PMID: 20192783.
104. Bokoch MP, Zou Y, Rasmussen SG, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi HJ, Thian FS, Kobilka TS, **Puglisi JD,** Weis WI, Pardo L, Prosser RS, Mueller L, Kobilka BK. (2010). Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature.* **463(7277)**:108-12. PMID: 20054398.
103. **Puglisi JD.** (2009). Resolving the elegant architecture of the ribosome. *Mol Cell.* **36(5)**:720-3. PMID: 20005832.
102. Elazar M, Liu M, McKenna SA, Liu P, Gehrig EA, **Puglisi JD,** Rossignol JF, Glenn JS. (2009). The anti-hepatitis C agent nitazoxanide induces phosphorylation of eukaryotic initiation factor 2alpha via protein kinase activated by double-stranded RNA activation. *Gastroenterology.* **137(5)**:1827-35. PMID: 19664635.
101. Marshall RA, Aitken CE, and **Puglisi JD.** (2009). GTP hydrolysis by IF2 guides progression of the ribosome into elongation. *Molecular Cell.* **35(1)**:37-47. PMID: 19595714.
100. Shimoike T, McKenna SA, Lindhout DA, **Puglisi JD.** (2009). Translational insensitivity to potent activation of PKR by HCV IRES RNA. *Antiviral Res.* **83(3)**:228-37. PMID: 19467267.
99. Marshall RA, Dorywalska M, **Puglisi JD.** (2008). Irreversible chemical steps control intersubunit dynamics during translation. *Proc Natl Acad Sci USA.* **105(40)**:15364-9. PMID: 18824686.
98. Uemura S, Iizuka R, Ueno T, Shimizu Y, Taguchi H, Ueda T, **Puglisi JD,** Funatsu T. (2008). Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic Acids Res.* **36(12)**:e70. PMID: 18511463.
97. **Puglisi JD,** Doudna JA. (2008). Nucleic acids and their protein partners. *Curr Opin Struct Biol.* **18(3)**:279-81. PMID: 18547801.
96. Aitken CE, Marshall RA, **Puglisi JD.** (2008). An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J.* **94(5)**:1826-35. PMID: 17921203.
95. Marshall RA, Aitken CE, Dorywalska M, **Puglisi JD.** (2008). Translation at the single-molecule level. *Annu Rev Biochem.* **77**:177-203. Review. PMID: 18518820.
94. **Puglisi JD.** (2007) Structural biology: The dance of domains. *Nature.* **450(7173)**:1171-2. PMID: 18097392.
93. McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, **Puglisi JD.** (2007) Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc.* **2(12)**:3270-7. PMID: 18079727.
92. Puglisi EV, **Puglisi JD.** (2007) Probing the conformation of human tRNA(3)(Lys) in solution by NMR. *FEBS Lett.* **581(27)**:5307-14. PMID: 17963705.
91. Gonzalez RL Jr, Chu S, **Puglisi JD.** (2007). Thiostrepton inhibition of tRNA delivery to the ribosome. *RNA.* **13(12)**:2091-7. PMID: 17951333.
90. Kim HD, **Puglisi JD,** Chu S. Fluctuations of Transfer RNAs between Classical and Hybrid States. *Biophys J.* **93(10)**:3575-82. PMID: 17693476.
89. Lee, TH, Blanchard, SC, Kim, HD, **Puglisi, JD,** Chu, S. (2007). The role of fluctuations in tRNA selection by the ribosome. *Proc Natl Acad Sci USA.* **104(34)**:13661-5.
88. McKenna SA, Lindhout DA, Shimoike T, **Puglisi JD.** (2007). Biophysical and biochemical investigations of dsRNA-activated kinase PKR. *Methods Enzymol.* Ch 14; Vol 430; pp. 373-396.

87. Marshall, R. A., **Puglisi JD.** (2007). NMR structural studies of aminoglycoside:RNA interaction. Ed. Arya, D.P. *Aminoglycoside Antibiotics*. John Wiley & Sons, pp. 181-207.
86. Lindhout DA, McKenna SA, Aitken CE, Liu CW, **Puglisi JD.** (2007). PKR: A NMR perspective. *JPNMRS*. **51**:199-215.
85. Varani L, Bankovich AJ, Liu CW, Colf LA, Jones LL, Kranz DM, **Puglisi JD**, Garcia KC. (2007). Solution mapping of T cell receptor docking footprints on peptide-MHC. *Proc. Nat'l Acad. Sci.* **104(32)**:13080-5.
84. McKenna SA, Lindhout DA, Shimoike T, Aitken CE, **Puglisi JD.** (2007). Viral dsRNA Inhibitors Prevent Self-association and Autophosphorylation of PKR. *J. Mol. Biol.* **372(1)**:103-13. PMID: 17619024.
83. Uemura S, Dorywalska M, Lee TH, Kim HD, **Puglisi JD**, Chu S. (2007). Peptide bond formation destabilizes Shine-Dalgarno interaction on the ribosome. *Nature*. **446(7134)**:454-7.
82. Sean A. McKenna, Darrin A. Lindhout, Colin Echeverría Aitken, and **Joseph D. Puglisi.** (2007). Chapter 8: Molecular Insights Into PKR Activation by Viral Double-Stranded RNA, **NATO Science Series**.
81. McKenna SA, Lindhout DA, Kim I, Liu CW, Gelev VM, Wagner G, **Puglisi JD.** (2007). Molecular framework for the activation of RNA-dependent protein kinase. *J. Biol. Chem.*, **282(15)**:11474-86.
80. Kim I, McKenna SA, Viani Puglisi E, **Puglisi JD.** (2007). Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA*. **13(2)**:289-94.
79. McKenna SA, Kim I., Liu CW, **Puglisi JD.** (2006). Uncoupling of RNA binding and PKR kinase activation by viral inhibitor RNAs. *J Mol Biol.* **358**:1270-85.
78. Kim I., Liu CW, **Puglisi JD.** (2006). Specific Recognition of HIV TAR RNA by the dsRNA Binding Domains (dsRBD1-dsRBD2) of PKR. *J Mol Biol.* **358**:430-42.
77. Lukavsky PJ, **Puglisi JD.** (2006). Structure Determination of Large Biological RNAs. *Methods Enzymol*; **394**:399-416.
76. Insil Kim, Peter Lukavsky, Geoff Otto, Corey Liu and **Joseph D. Puglisi.** (2005). Using NMR to Study large RNAs: Case Study of the HCV IRES. **IOS Press, NATO Science Series 364**:75-90.
75. Dorywalska M, Blanchard SC, Gonzalez RL, Kim HD, Chu S, **Puglisi JD.** (2005). Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic Acids Res.* **33(1)**:182-9
74. Underwood, K., Swartz, J. & **Puglisi JD.** (2005). Quantitative polysome analysis identifies limitation in cell-free protein synthesis. *Biotechnology and Engineering.* **91(4)**:425-35.
73. Otto GA and **Puglisi JD.** (2004). The pathway of HCV-IRES mediated translation initiation. *Cell.* **119(3)**:369-80.
72. Blanchard SC, Kim, H.D., Gonzalez RL, **Puglisi JD**, Chu, S. (2004). tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A*: **101(35)**:12893-8.
71. Blanchard SC, Gonzalez RL, Kim, H.D., Chu, S., **Puglisi JD.** (2004). tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol.* **11(10)**:1008-14.
70. **Puglisi JD** and Lukavsky PJ. (2004). Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* **10(5)**:889-93.
69. Runyon, S. and **Puglisi JD.** (2003). Design of a Cyclic Peptide that Targets a Viral RNA. *J. Am. Chem. Soc.* **125(51)**:15704-5.

68. Schroeder S.J., Fountain M.A., Kennedy SD, Lukavsky PJ, **Puglisi JD.** Krugh TR, Turner DH. (2003). Thermodynamic Stability and Structural Features of the J4/5 Loop in a *Pneumocystis carinii* Group I Intron. *Biochemistry* **42(48)**:14184-96.
67. Lukavsky P. J., Kim I., Otto GA, **Puglisi JD.** (2003). Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol.* **10(12)**:1033-8.
66. Li Q, Khosla C, **Puglisi JD,** and Liu CW. (2003). Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. *Biochemistry* **42(16)**:4648-57.
65. Lynch SR, Gonzalez RL, and **Puglisi JD.** (2003). Comparison of X-Ray Crystal Structure of the 30S Subunit-Antibiotic Complex with NMR Structure of Decoding Site Oligonucleotide-Paromomycin Complex. *Structure* **11(1)**:43-53.
64. Kim I., Lukavsky PJ, and **Puglisi JD.** (2002). NMR Study of 100 kDa HCV IRES RNA Using Segmental Isotope Labeling. *J. Am. Chem. Soc.* **124(32)**:9338-9.
63. Otto GA., Lukavsky PJ, Lancaster AM, Sarnow P. and **Puglisi JD.** (2002). Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction. *Proc. Nat'l Acad. Sci.* **8(7)**:913-23.
62. Yoshizawa S, Fourmy, D., Eason RG, and **Puglisi JD.** (2002). Sequence-Specific Recognition of the Major Groove of RNA by Deoxystreptamine. *Biochemistry.* **41**:6263-70.
61. Lukavsky PJ and **Puglisi JD.** (2001). RNAPack - An integrated NMR approach to RNA structure determination. *Met. Enz.* **25**:316-32.
60. Recht MI, and **Puglisi JD.** (2001). Aminoglycoside Resistance with Homogeneous and Heterogeneous Populations of Antibiotic-Resistant Ribosomes. *Antimicrob Agents Chemother.* **45(9)**:2414-9.
59. Blanchard SC and **Puglisi JD.** (2001). Solution structure of the A loop of 23S ribosomal RNA. *Proc. Nat'l Acad. Sci.* **98**:3720-25.
58. SR Lynch and **JD Puglisi** (2001). Structural Origins of Aminoglycoside Specificity for Prokaryotic Ribosomes *J. Mol. Biol.* **306**:1037-58.
57. SR Lynch and **JD Puglisi** (2001). Structure of a Eukaryotic Decoding Region A-site RNA *J. Mol. Biol.* **306**:1023-35.
56. PJ Lukavsky, GA Otto, AM Lancaster, P Sarnow and **JD Puglisi** (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nature Struct. Biol.* **7**:1105-10.
55. **Puglisi JD,** Blanchard SC, and Green R. (2000). Approaching translation at atomic resolution. *Nature Struct. Biol.* **7**:855-61.
54. Lynch SR and **Puglisi JD.** (2000). Application of Residual Dipolar Coupling Measurements to Identify Conformational Changes in RNA Induced by Antibiotics. *J. Am. Chem. Soc.,* **122**:7853-54.
53. **Puglisi JD.** (2000). mRNA processing: The 3'-end justifies the means. *Nature Struct. Biol.* **4**:263-4.
52. Dahlquist, K. and **Puglisi JD.** (2000). Interaction of Translation Initiation Factor IF1 with the E. coli ribosomal A site. *J. Mol. Biol.* **299**:1-15.
51. Green, R. and **Puglisi JD.** (1999). The ribosome revealed. *Nature Struct. Biol.* **6**:999-1004.
50. **Puglisi JD,** Blanchard, S. C., Dahlquist, K. D., Eason, R. G., Fourmy, D., Lynch, S., Recht, M. I., and Yoshizawa, S. (1999). Structural basis for aminoglycoside antibiotic action. In *Ribosomes99*, Chapter 34, pp. 419-29.

49. Yoshizawa, S., Fourmy, D., and **Puglisi JD.** (1999). RNA helix recognition by ribosomal RNA. *Science* **285**:1722-25.
48. Recht MI, Douthwaite S, and **Puglisi JD.** (1999). Basis for prokaryotic specificity of action of aminoglycoside antibiotics. *EMBO J.* **18**:3133-38.
47. Lynch SR, Recht MI, and **Puglisi JD.** (1999). Biochemical and NMR studies of aminoglycoside-RNA complexes. *Methods in Enzymology* **317**:241-61.
46. Recht MI, Douthwaite S, Dahlquist KD, and **Puglisi JD.** (1999). Effect of mutations in the A site of 16S rRNA on aminoglycoside antibiotic-ribosome interaction. *J. Mol. Biol.* **286**: 33-43.
45. **Puglisi JD** and Williamson, J.R. (1999). RNA interaction with small ligands and peptides. In *RNA World II* T. (Cech and R. Gesteland, Eds).pp. 403-25.
44. Puglisi, E. V. and **Puglisi JD.** (1998). Nuclear magnetic resonance spectroscopy of RNA. In *RNA Structure and Function* (R. W. Simons and M. Grunberg-Manago, Eds) pp. 117-46.
43. Puglisi, E. V. and **Puglisi JD.** (1998). HIV A-rich RNA loop mimics the tRNA anticodon structure. *Nature Struct. Biol.* **5**:1033-36.
42. Yoshizawa, S., Fourmy, D. and **Puglisi JD.** (1998). Structural origins of gentamicin antibiotic action. *EMBO J.* **17**:6437-48.
41. Yoshizawa, S. and **Puglisi JD.** (1998). NMR structure determination of an antibiotic-RNA complex. *NATO ASI series.* **301**:173-82.
40. Blanchard SC, Fourmy, D., Eason RG, and **Puglisi JD.** (1998). rRNA chemical groups required for aminoglycoside binding. *Biochemistry* **37**:7716-24.
39. Fourmy, D., Recht MI, and **Puglisi JD.** (1998). Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16S rRNA. *J. Mol. Biol.* **277**:347-62.
38. Fourmy, D., Yoshizawa, S. and **Puglisi JD.** (1998). Paromomycin binding induces a local conformational change in the A site of 16S rRNA. *J. Mol. Biol.* **277**:333-45.
37. **Puglisi JD.** (1997). Structural basis for aminoglycoside antibiotic action. In *Many Faces of RNA* pp. 97-111.
36. Puglisi, E. Viani and **Puglisi JD.** (1997). RNA structure. In *Post Transcriptional Gene Regulation* (D. Morris and J. Harford, Eds.) N.Y. Wiley and sons, pp. 1-22.
35. Puglisi, E.V., Noller, H. F., and **Puglisi JD.** (1997). Structure of a conserved RNA component of the peptidyl transferase center. *Nature Struct. Biol.* **4**:775-8.
34. Fourmy, D., Recht MI, Blanchard SC, and **Puglisi JD.** (1996). Structure of the A site of *E. coli* 16S rRNA complexed with an aminoglycoside antibiotic. *Science* **274**:1367-71.
33. Recht, M., Fourmy, D., Blanchard, S., Dahlquist, K., and **Puglisi JD.** (1996). RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J. Mol. Biol.* **262**:421-36.
32. **Puglisi JD.** (1995). Welcome guests to the RNA World: Proteins that interact with RNA. *Chemistry and Biology.* **2**:581.
31. **Puglisi JD,** Chen, L., Blanchard, S., and Frankel, A. (1995). Solution structure of an RNA-peptide complex. *Science.* **270(5239)**:1200-3.
30. **Puglisi JD** and Wyatt, JR (1995). Biochemical and NMR studies of RNA conformation. *Methods in Enzymology* **261**:323-50.

29. Puglisi, E.V., **Puglisi JD**, Williamson, J.R., and RajBhandary, U. L. (1994). NMR Analysis of tRNA acceptor stem microhelices: Discriminator base changes affect tRNA conformation at the 3'-end. *Proc. Nat'l Acad. Sci. USA*, **91**:11467-71.
28. **Puglisi JD**, Chen, L. Frankel, A.D., and Williamson, J.R. (1993). The Role of RNA Structure in Arginine Binding to TAR RNA. *Proc. Nat'l Acad. Sci. USA*, **90**:3680-4.
27. Giegé, R., Florentz, C., and **Puglisi JD**, (1993). tRNA Structure and Aminoacylation Efficiency. In *Prog. Nucl. Acids Res. Molec. Biol.* **45**, pp. 129-206.
26. Pütz, J., **Puglisi JD**, Florentz, C., and Giegé, R. (1993). Additive, Cooperative and Anticooperative Effects between Identity Nucleotides of a tRNA. *EMBO J.* **12**:2949-57.
25. **Puglisi JD**, Pütz, J., Florentz, C., and Giegé, R. (1993). Influence of tRNA Tertiary Structure and Stability on Aminoacylation by yeast Aspartyl-tRNA Synthetase. *Nucleic Acids Res.* **21**:41-49.
24. Batey, R.T., Inada, M., Kujawinski, E., **Puglisi JD**, and Williamson, J.R. (1992). Preparation of Isotopically Labeled Ribonucleotides for Multidimensional NMR Spectroscopy of RNA. *Nucleic Acids Res.* **20**:4515-23.
23. **Puglisi JD**, Tan, R., Calnan, B. J., Frankel, A. D. and Williamson, J.R. (1992). Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* **257**:76-80.
22. **Puglisi JD**, Tan, R., Frankel, A. D., Williamson, J. R. (1992). Structure and Function of HIV TAR RNA. In *Structural Tools for the Analysis of Protein-Nucleic Acid Complexes Advances in Life Sciences* (D. Lilley, H. Heumann, D. Suck, ed.) pp. 269-85.
21. Perret, V., Florentz, C., **Puglisi JD** and Giegé, R. (1992). Effect of Conformational Features on the Aminoacylation of tRNAs and Consequences on the Permutation of tRNA Specificities. *J. Mol. Biol.* **226**:323-33.
20. Rudinger, J., **Puglisi JD**, Pütz, J., Schatz, D., Eckstein, F., Florentz, C. and Giegé, R. (1992). Determinant Nucleotides of Yeast tRNA^{Asp} Interact Directly with Aspartyl-tRNA Synthetase. *Proc. Nat'l Acad. Sci. USA*, **89**:5882-6.
19. Wyatt, J. R., Chastain, M. and **Puglisi JD**. (1991). Synthesis and Purification of Large Amounts of RNA Oligonucleotides. *BioTechniques* **11**:764-9.
18. Pütz, J., **Puglisi JD**, Florentz, C. and Giegé, R. (1991). Identity Elements for Specific Aminoacylation of Yeast tRNA^{Asp} by Cognate Aspartyl-tRNA Synthetase. *Science*, **252**:1696-9.
17. **Puglisi JD**, Wyatt, J. R. and Tinoco, I., Jr. (1991). RNA Pseudoknots. *Accts Chem. Res.*, **24**:152-8.
16. Giegé, R., Florentz, C., Garcia, A., Grosjean, H., Perret, V., **Puglisi JD**, Theobald-Dietrich, A., and Ebel, J.-P. (1990). Exploring the Aminoacylation Function of transfer RNA by Macromolecular Engineering Approaches. Involvement of conformational features in the charging process of yeast tRNA^{Asp}. *Biochimie*, **72**:453-61.
15. Perret, V., Garcia, A., **Puglisi JD**, Grosjean, H., Ebel, J.-P. and Giegé, R. (1990). Conformation in Solution of tRNA^{Asp} Transcripts Deprived of Modified Nucleotides. *Biochimie* **72**:735-43.
14. Wyatt, J. R., **Puglisi JD** and Tinoco, I., Jr. (1990). RNA Pseudoknots: Stability and Loop Size Requirements. *J. Mol. Biol.* **214**, 455-70.
13. **Puglisi JD**, Wyatt, J. R. and Tinoco, I., Jr. (1990). Conformation of an RNA Pseudoknot. *J. Mol. Biol.* **214**:437-53.

12. Tinoco, I. Jr., **Puglisi JD** and Wyatt, J. R. (1990). RNA Folding. In Nucleic Acids and Molecular Biology, 4 (F. Eckstein and D. Lilley, eds.), Springer-Verlag, Berlin, pp. 205-26.
11. **Puglisi JD**, Wyatt, J. R. and Tinoco, I., Jr. (1990). Solution Conformation of an RNA Hairpin Loop. *Biochemistry* **29**:4215-26.
10. Wyatt, J. R., **Puglisi JD** and Tinoco, I. Jr., (1989). RNA Folding: Pseudoknots, Loops and Bulges. *Bioessays* **11**:100-6.
9. **Puglisi JD** and Tinoco, I., Jr. (1989). Absorbance Melting Curves of RNA. *Methods in Enzymology, RNA Processing 180* (J. E. Dahlberg and J. N. Abelson, eds), Academic Press, New York.
8. Tinoco, I., Jr., Aboul-ela, F., Hardin, C.C., **Puglisi JD**, Varani, G., Walker, G.T., Wolk, S. and Wyatt, J. R. (1989). Nucleic Acids from A to Z. In Frontiers of Macromolecular Science, (T. Saegusa, T. Higashimura and A. Abe, eds.) pp. 519-24, Blackwell Scientific Publications, Oxford.
7. **Puglisi JD**, Wyatt, J. R. and Tinoco, I., Jr. (1988). A Pseudoknotted RNA Oligonucleotide. *Nature*, **331**:283-6.
6. Wyatt, J. R., **Puglisi JD** and Tinoco, I., Jr. (1988). Pseudoknotted RNA Oligonucleotides. In Molecular Biology of RNA, (T. Cech, ed.) UCLA Symposia Series , **94**:25-32, Alan. R. Liss, Inc., New York, N.Y.
5. Tinoco, I. Jr., Davis, P.W., Hardin, C.C., **Puglisi JD**, Walker, G.T. and Wyatt, J. R. (1988). RNA Structure from A to Z. In Cold Spring Harbor Symp. Quant. Biol. **52**, pp 135-46.
4. Trulson, M.O., Cruz, P., **Puglisi JD**, Tinoco, I., Jr. and Mathies, R.A. (1987). Raman Spectroscopic Study of Left-Handed Z-RNA. *Biochemistry* **26**: 8624-30.
3. Hardin, C.C., Zarling, D.A., **Puglisi JD**, Trulson, M.O., Davis, P.D. and Tinoco, I. Jr. (1987). Stabilization of Z-RNA by chemical bromination and its recognition by anti-Z-DNA antibodies. *Biochemistry* **26**:5191-9.
2. Cruz, P., Hall, K., **Puglisi JD**, Davis, P., Hardin, C., Trulson, M., Mathies, R., Tinoco, I., Jr., Johnson, W. Jr. and Neilson, T. (1986). The Left-Handed Z-Form of Double-Stranded RNA. In Biomolecular Stereodynamics, IV, Adenine Press, New York, pp 179-200.
1. Tinoco, I. Jr., Cruz, P., Davis, P., Hall, K., Hardin, C.C., Mathies, R. A., **Puglisi JD**, Trulson, M.O., Johnson, W.C., and Neilson, T. (1986). Z-RNA: A Left-Handed Double Helix. In Structure and Dynamics of RNA, Plenum Press, New York, pp. 55-68.

Books:

- Biophysics and Structure to Counter Threats and Challenges. NATO Science for Peace and Security Series B: Physics and Biophysics. Puglisi, Joseph D. **(Ed.)** (2013). Publisher: Springer.
- Biophysics and the Challenges of Emerging Threats. NATO Science for Peace and Security Series. Puglisi, Joseph D. **(Ed.)** (2009). Publisher: Springer.
- Structure and Biophysics: New Technologies for Current Challenges in Biology and Beyond. NATO Science for Peace and Security Series. Puglisi, Joseph D. **(Ed.)**, (2007). Publisher: Springer.
- Structure, Dynamics and Function of Biological Macromolecules and Assemblies. **Volume 364** NATO Science Series: Life and Behavioural Sciences. Puglisi, Joseph D. **(Ed.)** (2005). Publisher: IOS Press.
- Physical Chemistry: Principles and Applications in Biological Sciences (4th Edition) Ignacio Tinoco, Kenneth Sauer, James C. Wang, and **Joseph D. Puglisi**, 06 August 2001.

Invited Presentations: More than 300.