

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GUARDANT HEALTH, INC.
Petitioner,

v.

COLD SPRING HARBOR LABORATORY
Patent Owner

IPR2025-XXXXXX
U.S. PATENT NO. 10,947,589

DECLARATION OF GREGORY COOPER, PH.D IN SUPPORT OF
PETITIONER FOR *INTER PARTES* REVIEW

Mail Stop "PATENT BOARD"
Patent Trial and Appeal Board
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. QUALIFICATIONS AND COMPENSATION.....	1
III. SUMMARY OF OPINIONS.....	6
IV. LEGAL STANDARDS	7
A. Claim Construction	8
B. Obviousness.....	8
V. THE '589 PATENT.....	9
VI. PROSECUTION HISTORY OF THE '589 PATENT.....	14
VII. RECENT BOARD UNPATENTABILITY DECISION IN IPR PROCEEDINGS BETWEEN PARSE BIOSCIENCES AND 10X GENOMICS	15
VIII. CLAIM CONSTRUCTION	18
IX. PERSON OF ORDINARY SKILL IN THE ART	18
X. TECHNICAL BACKGROUND AND STATE OF THE ART	19
A. DNA	19
B. Amplification.....	21
C. DNA Tagging	23
D. Amplification Bias and Molecular Tagging.....	27
E. Sequencing	30
F. Copy Number	32
XI. THE PRIOR ART REFERENCES OF THE GROUNDS	33
A. Lo (Ex. 1010)	33
1. Lo is § 102 Prior Art	33
2. Lo Overview	33
B. McCloskey (Ex. 1008)	41
1. McCloskey is § 102 Prior Art	42
2. McCloskey Overview	42

TABLE OF CONTENTS

(continued)

	Page
C. Brenner (Ex. 1014).....	45
1. Brenner is § 102 Prior Art.....	45
2. Brenner Overview	45
D. Chee (Ex. 1015).....	49
1. Chee is § 102(e) Prior Art.....	49
2. Chee Overview.....	50
E. Porreca (Ex. 1017).....	54
1. Porreca is § 102(e) Prior Art.....	54
2. Porreca Overview.....	54
F. Quake (Ex. 1019)	58
1. Quake is Prior Art	58
2. Quake Overview	58
XII. PRIOR ART RENDERS OBVIOUS THE CLAIMS OF THE '589 PATENT	62
A. Ground 1: Claims 1-3, 6-8, 11, and 16-18 Are Unpatentable as Obvious Over Lo in View Of The Knowledge of a POSA As Reflected In McCloskey, Brenner, Chee, And Porreca	63
1. Claim 1:.....	63
a) [1A] A method comprising: a) obtaining segments of genomic nucleic acids from a sample;	63
b) [1B] tagging the segments with nucleic acid tags, thereby generating a plurality of unique tagged nucleic acid molecules, such that each of the unique tagged nucleic acid molecules comprises a segment of the segments of the genomic nucleic acids from step (a) and a nucleic acid tag of the nucleic acid tags;.....	64

TABLE OF CONTENTS

(continued)

	Page
c) [1C] subjecting the plurality of the unique tagged nucleic acid molecules to a polymerase chain reaction (PCR), thereby generating amplified tagged nucleic acid molecules;	78
d) [1D] generating tag associated sequence reads by sequencing the amplified tagged nucleic acid molecules of step (c);.....	85
e) [1E] assigning each of the amplified tagged nucleic acid molecules to a location on a reference genome from the same species from which the sample has been obtained by mapping each of the tag associated sequence read of step (d) to a location of the reference genome; and.....	88
f) [1F] at a plurality of locations on the reference genome, counting the number of unique tagged nucleic acid molecules that have been assigned to the same location on the reference genome based on the mapping step to obtain a count for each location of the plurality of locations on the reference genome, thereby obtaining copy number information from genomic nucleic acids in the sample.	91
2. Claim 2: The method of claim 1, further comprising estimating a copy number at multiple locations within a region of the reference genome.....	95
3. Claim 3: The method of claim 1, further comprising comparing a count from counts obtained in step (f) for a location on the reference genome to a count for the same location on the reference genome obtained from a reference sample, thereby estimating a relative copy number at the location.....	101

TABLE OF CONTENTS

(continued)

	Page
4. Claim 6: The method of claim 1, further comprising (l) summing the counts from the counts obtained in step (f) for locations at a region of the reference genome; (m) summing the counts obtained from a reference sample for the same region of the reference genome; and comparing the sums of the counts obtained in step (l) and step (m), thereby estimating a relative copy number at the region of the reference genome.....	103
5. Claim 7: The method of claim 1, wherein the unique tagged nucleic acid molecules are subject to a hybrid capture method prior to PCR in step (c) or the amplified tagged nucleic acid molecules produced in step (c) are subject to a hybrid capture method prior to step (d).....	105
6. Claim 8: The method of claim 7, wherein each of the unique tagged nucleic acid molecules subjected to the hybrid capture method has a nucleic acid sequence which is complementary to a capture probe on an array or on a microarray.	106
7. Claim 11: The method of claim 1, wherein each of the tagged nucleic acid molecules comprise more than one tag.....	107
8. Claim 17: The method of claim 1, wherein the genomic nucleic acids are obtained from blood or other bodily fluids.....	111
9. Claim 18: The method of claim 1, wherein the genomic nucleic acids are obtained from tumor cell DNA in a bodily fluid.....	112
B. Ground 2: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In Brenner Or McCloskey Renders Obvious Claim 9.....	113
1. Claim 9: The method of claim 1, wherein each of the tagged nucleic acid molecules differs at more than one nucleotide.	113

TABLE OF CONTENTS

(continued)

	Page
C. Ground 3: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In McCloskey, Brenner, or Porreca Renders Obvious Claims 12 And 14-16.....	116
1. Claim 12: The method of claim 1, wherein the tags further comprise a sample tag.	116
2. Claim 14: The method of claim 1, wherein the unique tagged nucleic acid molecules are pooled with another plurality of unique tagged nucleic acid molecules having a different sample tag prior to the PCR or prior to said sequencing of the amplified tagged nucleic acid molecules of step (c).	119
3. Claim 15: The method of claim 14, further comprising deconvoluting the tag associated sequence reads by grouping the tag associated sequence reads.....	120
4. Claim 16: The method of claim 1, further comprising obtaining copy number information from genomic nucleic acids of different species in a population using the method of claim 1, and determining a relative count of the genomic nucleic acids of the different species in a population by comparing the copy number information obtained from the genomic nucleic acids of the different species.	120
D. Ground 4: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In Chee or Porreca Renders Obvious Claim 13.....	122
1. Claim 13: The method of claim 1, wherein the tagged nucleic acid molecules generated from step (b) and that differ by a single nucleotide are ignored during or after the counting step.	122
E. Ground 5: Claims 4 and 5 Are Unpatentable as Obvious Over Quake in View Of McCloskey, Brenner, Chee, or Porreca and Further in View of the Knowledge of a POSA	126

TABLE OF CONTENTS

(continued)

Page

1.	Claim 4: The method of claim 1, further comprising g) summing counts obtained in step (f) for locations on a first region of the reference genome, wherein the first region comprises more than one location; h) summing counts obtained in step (f) for locations on a second region of the reference genome, wherein the second region comprises a number of locations which is comparable to the number of locations of the first region; i) comparing the counts obtained in step (g) to the counts obtained in step (h), thereby estimating a relative copy number at the first region of the reference genome to the copy number at the second region of the reference genome.	126
2.	Claim 5: The method of claim 4, wherein step (h) further comprises j) summing counts obtained in step (f) for locations at a third region of the reference genome, wherein the third region comprises a number of locations which is comparable to the number of locations of the first region; and k) obtaining an average of the sum of the counts obtained in step (f) for locations at the second region and the sum of the counts obtained in step (j) for locations at the third region.....	133
F.	Ground 6: Claims 10 is Unpatentable as Obvious Over Lo In View Of The Knowledge Of A POSA In Further View Of Specific Disclosure In Chee, Porreca, Or Brenner.....	135
1.	Claim 10: The method of claim 1, wherein said tagging the segments with nucleic acid tags comprises ligation of adaptors comprising the tags to at least one end of the segments of the genomic nucleic acids.....	135
DECLARATION		139

I. INTRODUCTION

1. My name is Greg Cooper. I have been retained by Guardant Health, Inc. (“Guardant” or “Petitioner”) in the above-captioned *inter partes* review (“IPR”) to provide my independent evaluation and opinions regarding U.S. Patent No. 10,947,589 (the “’589 patent”) Ex. 1001 and the related prior art.

2. This Declaration sets forth my analyses and opinions based on my technical background, skills, knowledge, training, education and experience in the relevant art, and the materials I considered. As I explain below, it is my opinion that claims 1-18 of the ’589 patent are rendered obvious by the combination of prior art references discussed herein.

3. In parallel with the analysis of the ’589 patent, I have also been retained by Guardant to provide an evaluation of the related patent U.S. Patent No. 12,234,510 (“’510 patent”). My analysis of the ’510 patent is very similar to the instant ’589 patent given the analogous claim language of the ’510, and thus I reserve the right to rely on that declaration in support of my opinions of the ’589 patent herein.

II. QUALIFICATIONS AND COMPENSATION

4. I am currently a Faculty Investigator at the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama. I have held this position since 2010. HudsonAlpha, established in 2008, is an independent, non-profit institution

dedicated to improving human health and welfare through biological research, genomic medicine, education, and economic development. Since coming to HudsonAlpha, I have set up an independent human genomics research lab. My lab works on projects related to the genetics of human disease, with a particular emphasis on research and clinical applications of genome sequencing for diagnosing children with developmental delays and intellectual disabilities. We also study the genetics and biochemistry of gene expression regulation, develop new computational approaches for analysis of human genomes and related datasets, and pursue other applications of genomic technologies to understand human biology.

5. I am also concurrently an Adjunct Associate Professor in the Department of Genetics at the University of Alabama at Birmingham. I have held an adjunct faculty position since 2010.

6. I earned a Bachelor of Arts in Microbiology and Bachelor of Science in Mathematics and Statistics, summa cum laude, from Miami University (Oxford, OH) in 2001. As an undergraduate student, I performed molecular biology and genetic research in research laboratories at Miami University, the University of Illinois at Urbana-Champaign, and the University of Cincinnati.

7. I earned a Ph.D. in Genetics from Stanford University in 2006 under the mentorship of Professor Arend Sidow, Ph.D. My thesis work was mainly comprised of computational biology research, with a focus on evolutionary

comparisons of mammalian genome assemblies to better annotate functional elements in human genomes and assess whether mutations in those elements might contribute to disease. While there, I worked and/or received training in areas including: theoretical and applied human genetics, molecular evolution, functional genomics, developmental biology, bioinformatics, genomic technologies, and analysis of datasets produced by high-throughput experimental platforms, especially DNA sequencers.

8. From 2006 to 2009, I was a Senior Fellow in the Department of Genome Sciences at the University of Washington working with Professors Evan Eichler, Ph.D., and Deborah Nickerson, Ph.D. While there, I conducted my postdoctoral research. I was primarily interested in applying genomic technologies to better understand the genetic basis for human disease. One series of projects centered on analytical and algorithmic developments to identify large deletion and duplication events, also known as “copy-number variants” (“CNVs”), from microarray data generated to perform Single Nucleotide Polymorphism (“SNP”) genotyping. This process involved extensive processing, normalization, and analysis of Illumina SNP array data. A second series of projects was aimed at better understanding the genetic basis for a variety of complex traits. These included genetic analyses of transcript levels, for example, and also a genome-wide association study of patient response to the commonly prescribed anticoagulant

warfarin. Finally, another series of projects focused on analysis of high-throughput DNA-sequencing data, in particular the locations of genetic variants, from individuals with Mendelian diseases. I completed my postdoctoral studies in 2009.

9. From 2009 to 2010, I was an Acting Assistant Professor in the Department of Genome Sciences at the University of Washington.

10. I am an author or contributor to more than 100 publications in the areas of human genetics and genomics, including functional and evolutionary annotations of mammalian genomes, DNA sequencing, and DNA analysis, as they relate to human diseases, especially, but not exclusively, intellectual disabilities and other developmental issues. As of June 23, 2025, these publications have been collectively cited more than 50,000 times, according to Google Scholar estimates.

11. Among the awards and honors I have earned are: Harrison Scholar, Miami University (Oxford, OH) (1997-2001), Goldwater Scholar (1999-2001), Howard Hughes Medical Institute Doctoral Fellowship (2001-2006), a University of Washington Genome Training Grant Postdoctoral Fellowship (2007), and a Merck, Jane Coffin Childs Memorial Fund Postdoctoral Fellowship (2007-2010).

12. I am a Principal Investigator (“PI”) on a current human genomic research project funded by the National Institutes of Health (“NIH”) and a PI or co-Investigator on several additional projects funded by non-NIH sources, including both governmental and philanthropic sources. I have also been a PI or co-

Investigator on multiple additional grants related to human genetics and genomics, from NIH and elsewhere, that have been completed.

13. In general, I have knowledge and experience with both the general scientific contexts relevant to this matter and specific details of the technologies employed. For example, as described here, I have extensive knowledge about molecular techniques such as DNA sequencing, molecular tagging, PCR, gene expression analysis, alignment and counting of sequencing reads, and detection of genetic variation such as changes to copy number.

14. I have also previously served as an expert witness in three now concluded patent disputes and two ongoing patent disputes. The three completed cases, in which I worked on behalf of the plaintiffs, are Illumina/Verinata vs Ariosa/Roche (C.A. No. 3:12-5501-SI, N.D. Cal.), Guardant Health vs FMI (C.A. No. 17-1616-LPS-CJB, D. Del.), and Guardant Health vs PGDx (17-1623-LPS-CJB, D. Del.). The two ongoing cases, in which I am working on behalf of the defendants, are Natera vs ArcherDX/Invitae (C.A. No. 20-125-GBW, D. Del.) and 10X Genomics v. Parse Bioscience (C.A. No. 22-1117-JHS, D. Del.). I have also recently agreed to serve as an expert on behalf of Guardant as the plaintiff in an infringement suit against Tempus AI, Inc. (1:24-cv-00687, GBW, D. Del), although I have not yet begun work on that case as of July 8, 2025.

15. A copy of my Curriculum Vitae, attached as Exhibit 1023, contains further details on my education, experience, publications, and other qualifications to render an expert opinion in this matter.

16. For this service, I am paid my standard hourly consulting fee of \$750/hour, and I am reimbursed for actual expenses. My compensation in no way depends on the outcome of this matter or the substance of this declaration.

III. SUMMARY OF OPINIONS

17. Based on my knowledge, experience, and the materials that I have reviewed, it is my opinion that claims 1-3, 6-8, 11, and 17-18 of the '589 patent are obvious based on U.S. Patent Application No. 2009/0029377 to Lo et al. ("Lo") (Ex. 1010) in view of the knowledge of the skilled artisan regarding molecular counting by tagging, amplification, and sequencing. This knowledge of the skilled artisan is reflected in at least four references: (1) M. McCloskey et al., *Encoding PCR Products with Batch-stamps and Barcodes*, Biochemical Genetics, Vol. 45 (Dec. 2007) ("McCloskey") (Ex. 1008), (2) U.S. Patent No. 7,537,897 to Brenner et al. ("Brenner") (Ex. 1014), (3) U.S. Patent No. 9,085,798 to Chee ("Chee") (Ex. 1015), and (4) U.S. Patent No. 11,840,730 to Porreca et al. ("Porreca") (Ex. 1017).

18. It is also my opinion that the following claims are obvious based on the following combinations:

- Claims 4-5 are obvious based on Lo and the knowledge of the skilled artisan in further view of Quake.
- Claim 9 is obvious based on Lo and the knowledge of the skilled artisan in further view of specific disclosure in McCloskey and Brenner.
- Claims 12 and 14-16 are obvious based on Lo and the knowledge of the skilled artisan in further view of specific disclosure in McCloskey, Brenner, or Porreca.
- Claim 13 is obvious based on Lo and the knowledge of the skilled artisan in further view of disclosure in Chee or Porreca.
- Claim 10 is obvious based on Lo and the knowledge of the skilled artisan in further view of disclosure in Chee, Porreca or Brenner.

IV. LEGAL STANDARDS

19. I am not a lawyer and offer no legal opinion in this declaration. I have, however, been advised by counsel as to various legal standards that apply to the technical issues I address in this declaration. I have applied those standards in arriving at my conclusions expressed in this declaration. The recitation of the legal

standards set forth here is virtually identical to those I relied upon in a recent set of IPR proceedings where I testified on behalf of Parse Biosciences.

A. Claim Construction

20. I have been informed that the words of the claims of a patent are generally given the ordinary and customary meaning as interpreted by a person of ordinary skill in the art (POSA) at the time of the invention. I have been informed that patent claims are construed in light of the claim language, the patent specification (including the drawings), and the prosecution history (including the references cited during the prosecution of the patent).

B. Obviousness

21. I understand that even if a patent is not anticipated, a claimed invention is “obvious” if the difference between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a POSA.

22. I understand that a POSA provides a reference point from which the prior art and claimed invention should be viewed. This reference point prevents the use of one’s insight or hindsight in deciding whether a claim is obvious.

23. Also, I understand that a claim can be obvious in light of a combination of multiple prior art references. I understand that such a combination must meet certain requirements. I understand that the prior art must teach or suggest all the

requirements of the invention as recited in the claim. I understand that the prior art, coupled with knowledge generally available in the art at the time of invention, must contain some suggestion or incentive that would have motivated a POSA to modify a reference or to combine the references. I also understand that this proposed modification must have had a reasonable expectation of success.

24. I understand that only analogous art may be considered in an obviousness analysis. Art is considered analogous if it is in the same field of endeavor as the patent or if it is reasonably pertinent to the particular problem with which the patent is involved.

25. I also understand that in evaluating whether patent claims are obvious, objective indicia of non-obviousness can be considered, such as but not limited to, commercial success, long-felt but unsolved needs, praise, unexpected results, industry acceptance, failure of others, and skepticism by experts.

V. THE '589 PATENT

26. The '589 patent is titled "Varietal Counting of Nucleic Acids For Obtaining Genomic Copy Number Information." *See generally* Ex. 1001. The claims of the '589 patent generally recite methods for determining genomic copy number information by tagging nucleic acids segments, making copies of the segments, sequencing the segments, and counting the number of unique tagged molecules at genome locations.

27. Claim 1 is representative and is reproduced below:

1. A method comprising:

- (a) obtaining segments of genomic nucleic acids from a sample;
- (b) tagging the segments with nucleic acid tags, thereby generating a plurality of unique tagged nucleic acid molecules, such that each of the unique tagged nucleic acid molecules comprises a segment of the segments of the genomic nucleic acids from step (a) and a nucleic acid tag of the nucleic acid tags;
- (c) subjecting the plurality of the unique tagged nucleic acid molecules to a polymerase chain reaction (PCR), thereby generating amplified tagged nucleic acid molecules;
- (d) generating tag associated sequence reads by sequencing the amplified tagged nucleic acid molecules of step (c);
- (e) assigning each of the amplified tagged nucleic acid molecules to a location on a reference genome from the same species from which the sample has been obtained by mapping each of the tag associated sequence reads of step (d) to a location of the reference genome; and
- (f) at a plurality of locations on the reference genome, counting the number of unique tagged nucleic acid molecules that have been assigned to the same location on the reference genome based on the mapping step to obtain a count for each location of the plurality of locations on the reference genome,

thereby obtaining copy number information from genomic nucleic acids in the sample.

28. The background of the '589 patent explains that genomic copy number information was commonly obtained using whole genome amplification. *Id.* at 1:37-40. Whole genome amplification, like the name suggests, relies on amplifying, or making copies of, the entire genome or substantially all of the entire genome of a cell, as opposed to specific target regions. The background notes that the whole genome amplification method oversamples certain regions and introduces distortions. Ligation-mediated PCR solved some issues with whole genome analysis but was supposedly still susceptible to amplification bias. *Id.* at 1:37-40, 51-55, 60-66. The '589 patent purports to address amplification bias through the well-known use of tagging, amplification, and sequencing to perform molecular counting.

29. Consistent with this, the specification of the '589 patent reflects that the use of tagging before sequencing for molecular counting as recited in the claims is not inventive but rather a prior art technique. The specification explains, for instance, that “Miner et al. (2004) describe a method of molecular barcoding to label template DNA prior to PCR amplification, [which] allows for the identification of contaminant and redundant sequences by counting on distinctly tagged sequences.” *Id.* at 2:12-16. Similarly, the specification notes that McCloskey et al. (2007) describes a method of barcoding “template DNA molecules prior to PCR

amplification” and that U.S. Patent No. 7,537,897 (Brenner) describes “methods for molecular counting by labeling molecules of an input sample with unique oligonucleotide tags and subsequently amplifying and counting the number of different tags.” *Id.* at 2:16-20, 23-26.

30. The specification, in my opinion, offers no satisfactory explanation for how these prior methods are materially different from the alleged invention. McCloskey, for instance, is criticized as using “template specific primers to barcode template DNA molecules prior to PCR amplification,” stating that this approach has an alleged drawback of requiring primers for each species of template molecule that is studied. *See id.* at 2:23-29. Yet, nothing in the claims excludes this approach. To the contrary the patent clearly states that the tags can be introduced by a range of methods including “priming.” For example, the specification states, “[t]ags may be associated with nucleic acid molecules by using, singly or in combination, ligation, nick-ligation, terminal transferase, hybridization, priming, and the like.” *Id.* at 21:38-40. Thus, it cannot be the case that McCloskey are distinguishable from the alleged invention on the basis of introducing tags via priming. The ’589 patent’s alleged distinction over Miner and Brenner is even more peculiar. Specifically, the patent purports to criticize these references because they use ligation for tagging, which is supposedly an “inefficient” reaction. *See id.* at 2:20-22. This is peculiar because not only does the patent have an entire section on how to perform “Tagging

by Ligation,” *see id.* at 22:43-24:29, but explicitly claims this process in claim 10. Thus, it simply cannot be the case that Miner and Brenner are distinguishable from the alleged invention on the basis of their use of ligation for tagging. In general, the claims of the ’589 patent recite prior art techniques for adding tags, and neither the specification nor the claims offer any explanation or basis for distinguishing over those prior art tagging methods.

31. The specification of the ’589 patent further discloses that PCR was an amplification method known in the art. Specifically, the specification explains that “[g]enerally...the [amplification] steps are performed multiple times (e.g., polymerase chain reaction).” *Id.* at 13:50-54. The Abstract also references “subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction” and Claim 1 recites “subjecting the...tagged nucleic acid molecules to a polymerase chain reaction (PCR), thereby generating amplified tagged nucleic acid molecules.” *Id.* at Abstract, Claim 1.

32. The specification of the ’589 patent discloses that high-throughput DNA sequencing was also well-known in the art. Specifically, the specification notes that any method directed towards determining an accurate copy number “should also be robust using existing methodologies for high volume, massively parallel sequencing.” *Id.* at 2:37-39. Nowhere does the specification suggest that massively parallel sequencing was something new or inventive.

33. I do not have an opinion on the priority date of the '589 patent. Based on the face of the patent, I understand that the earliest possible priority date of the '589 patent is October 22, 2010, based on the filing date of U.S. Provisional Patent Application No. 61/406,067. I have considered October 22, 2010 to be the relevant date for assessing prior art invalidity, despite the differences in claimed subject matter between the provisional application and the '589 patent.

VI. PROSECUTION HISTORY OF THE '589 PATENT

34. On March 7, 2016, Patent Owner filed Nonprovisional Patent Application No. 15/063,278 titled "Varietal Counting of Nucleic Acids for Obtaining Genomic Copy Number Information." Ex. 1009.140. I have been informed by counsel that during the prosecution of the '589 patent, the Examiner issued a non-final rejection on subject matter eligibility and indefiniteness grounds and a final rejection on subject matter eligibility and indefiniteness grounds. The Patent Owner overcame the final rejection following a meeting with the Examiner during which the Patent Owner agreed to amend the claim language. The Examiner also withdrew the subject matter eligibility grounds for rejection. I have been further informed by counsel that the Examiner issued the patent after the Patent Owner filed a terminal disclaimer subject to the expiration dates of U.S. 9,404,156 (the '156 Patent).

35. On March 16, 2021, the '589 patent was issued. Ex. 1001. In the Reasons for Allowance, the Examiner stated “No prior art either alone or in combination with the other art in the record teaches/teach or reasonably suggests/suggest a method which comprises all limitations recited in claims 88 and 94.” Ex. 1009.644.

36. During the prosecution of the '589 patent, the Patent Owner identified Brenner (Ex. 1014) and McCloskey (Ex. 1008) in an IDS. Ex. 1009.207, .209. The Examiner did not cite to or discuss Brenner or McCloskey in any of the non-final or final rejections. As discussed below, §§ XI.B.2, XI.C.2 *infra*, Brenner and McCloskey disclose the DNA tagging, amplification, sequencing, and counting limitations of claim 1 of the '589 patent. Lo (Ex. 1010) was not presented to the Examiner. Lo discloses the concept of obtaining DNA segments from a sample and “assigning each of the amplified tagged nucleic acid molecules to a location on a reference genome” to determine copy number information as recited by claim 1 of the '589 patent.

VII. RECENT BOARD UNPATENTABILITY DECISION IN IPR PROCEEDINGS BETWEEN PARSE BIOSCIENCES AND 10X GENOMICS

37. I understand that this is not the first time that the Board has reviewed a patent directed towards nucleic acid tagging and counting. In fact, I provided testimony in IPR proceedings between Parse Biosciences, Inc. and 10x Genomics,

Inc. that involved U.S. Patent No. 10,697,013 to Brenner *et.al.* (the “ ‘013 patent”) related to the use of tagging, amplification, and sequencing for molecular counting, which even relied on one of the same prior art references as the instant petition, namely the McCloskey reference. The specification of the ’013 patent explains, for instance, that “MIDs are employed to uniquely tag each individual polynucleotide in a sample. Identification of the number of unique MIDs in a sample can provide a readout of how many individual polynucleotides are present in the sample.” Ex. 1024 at 6:53-57.

38. Specifically, on February 5, 2024, the Board granted institution of *inter partes* review of the ’013 patent in proceeding IPR2023-00958. Ex. 1025 at 2. The ’013 patent claimed priority to a provisional application filed on August 20, 2009, and this priority date was not contested during the IPR proceedings involving the ’013 patent. The key point is that this priority date is well before the priority date of the patent at issue here. Therefore, to the extent the ’013 patent that was at issue in IPR2023-00958 is invalid, it is my opinion that the ’589 patent would be similarly invalid, as I explain further below.

39. Indeed, claim 1 of the ’013 patent contains many of the same elements as the ’589 patent. Step (a) consists of “providing a sample comprising ... a plurality of polynucleotides.” Step (b) consists of “performing combinatorial tagging to generate a plurality of tagged polynucleotides.” This tagging process in step (b)

involves introducing a “second tag sequence” that is used for “distinguishing said sample polynucleotide from other sample polynucleotides from said cell.” Step (c) consists of “amplifying said tagged polynucleotides.” Step (d) consists of “sequencing said plurality of amplified polynucleotides.” Finally, step (e) consists of “using the sequences in step (d) to count sample polynucleotides.” Thus, claim 1 of the ’013 is directed to the same general counting process as claimed in the ’589 patent involving tagging, amplification, and sequencing. *See generally* Ex. 1024.

40. After a full review of the merits, the Board found that all claims were unpatentable as obvious in light of prior art. In its final written decision issued on February 3, 2025, claims 1-12 and 19-28 were found unpatentable as obvious using Linnarsson (Ex. 1026) in combination with McCloskey (Ex. 1008) and further in view of the knowledge of a POSA. Similarly, claims 13-18 were found unpatentable as obvious using Linnarsson (Ex. 1026) in combination with McCloskey (Ex. 1008) and a second McCloskey reference (Ex. 1027)) and further in view of the knowledge of a POSA. Ex. 1005 at 51.

41. As noted above, I submitted testimony in support of the petitioner in the IPR proceedings. *See* Ex. 1028. I reserve the right to rely upon my opinions in these prior proceedings here. Given the Board’s decision to invalidate the ’013 patent in the previous proceeding where I testified, it is difficult to see how the ’589

patent claims could not be invalid, particularly when one considers that the '013 patent has an earlier priority date.

VIII. CLAIM CONSTRUCTION

42. I understand that for purposes of this Petition, the claims will be interpreted consistent with the Patent Owner's arguments and positions in the parallel district court proceeding, particularly in view of its contentions regarding how the '589 patent is allegedly infringed. Ex. 1020.

IX. PERSON OF ORDINARY SKILL IN THE ART

43. I understand that whether a patent is obvious is analyzed from the perspective of a POSA at the time of the alleged invention. I am knowledgeable about the level of ordinary skill in the art of the time of the alleged invention based on my education and experience. *See supra* § II.

44. In my opinion, a POSA would have been someone with a Master's degree and/or Ph.D. in molecular biology, genetics, chemistry, engineering or a related discipline and at least two years of post-Master's, postdoctoral, or industry experience. Alternatively, a POSA would have been someone with a Bachelor of Science degree in molecular biology, genetics, chemistry, engineering or a related discipline and at least five years of industry or academic experience (including any experimental work toward a graduate degree) relating to DNA labeling techniques, DNA amplification methods, and DNA sequencing technologies.

45. Further, it is my opinion that a POSA would have been familiar with associated tools, methods, and techniques including: (1) DNA tagging, (§ X.C); (2) nucleic acid amplification, (§ X.B); (3) use of tags to overcome amplification bias, (§ X.D); and (4) sequencing; alignment, and counting of sequencing reads (§ X.E).

X. TECHNICAL BACKGROUND AND STATE OF THE ART

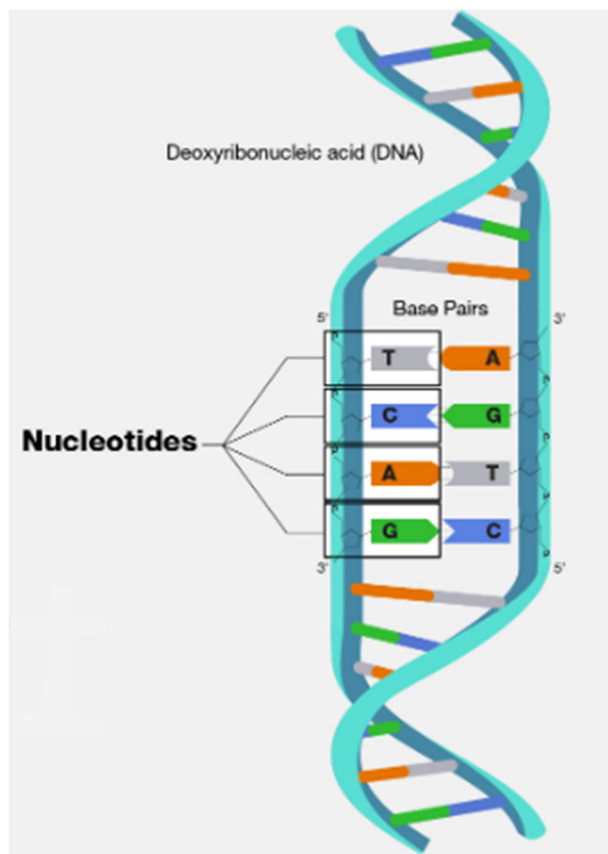
46. Below, I provide an overview of the relevant technology and the state of the art. Much of this background discussion is similar to that from my prior testimony on behalf of Parse Biosciences.

A. DNA

47. DNA, deoxyribonucleic acid, is a biological molecule that contains the information needed for cellular life. Individual cells use DNA for many aspects of cell development and maintenance. This information is stored in the sequence of adjacent bases in the DNA strand, or a “nucleotide sequence.” DNA is one type of “nucleic acid” and is thus commonly referred to as such.

48. DNA is a double-stranded molecule composed of units called nucleotides. A polynucleotide is a linear polymer comprised of multiple nucleotides. An oligonucleotide, or “oligo,” is a short polynucleotide. Oligonucleotides are commonly used, *inter alia*, as primers with various biological techniques such as DNA sequencing, polymerase chain reaction (“PCR”), and molecular cloning.

49. In DNA, each nucleotide comprises a deoxyribose sugar molecule, one of four nitrogenous bases, and one or more phosphate groups. The four nitrogenous bases of DNA are: adenine (“A”), thymine (“T”); cytosine (“C”), and guanine (“G”). The two strands of DNA are connected through complementary base pairing via chemical bonds, also known as “hybridization.” A pairs with T, and C pairs with G, as depicted below:



Ex. 1029 at 2

50. Each strand of the DNA molecule has two ends – one end is referred to as 5' and the other end is referred to as 3'. DNA is synthesized in the 5' to 3' direction.

B. Amplification

51. Before it can be sequenced, a sample with a limited amount of DNA must sometimes be subject to amplification. Amplification refers to the process of creating multiple copies of one or more polynucleotides present in an input sample, and can be applied to the whole genome or to fragments depending on the specific technique employed. Various methods of amplification were known in the art before 2011, such as polymerase chain reaction (“PCR”), ligase chain reaction (“LCR”), nucleic acid based amplification (“NASBA”), and transcription-mediated amplification, etc. *See, e.g.*, Ex. 1008 at Abstract (“Polymerase chain reaction (PCR) has become the mainstay of DNA sequence analysis”).

52. Brenner and McCloskey each describe amplification steps that were known and widely practiced in the art. *See* Ex. 1014 at 33:6-27; Ex. 1008 at 764-767. McCloskey, Chee, and Lo all likewise recognized PCR reactions as a conventional form of amplification. Ex. 1008 at Abstract (“Polymerase chain reaction (PCR) has become the mainstay of DNA sequence analysis”); Ex. 1015 at 3:2-3; Ex. 1010 ¶ 36. Indeed, the ’589 patent also acknowledges the use of the conventional amplification technique of PCR, described in detail below, as a part of its claims and description. Ex. 1001 at Abstract, Claim 1.

53. PCR is a technique that uses repeated cycles of DNA synthesis to amplify target nucleic acid sequences. Each cycle roughly doubles the amount of

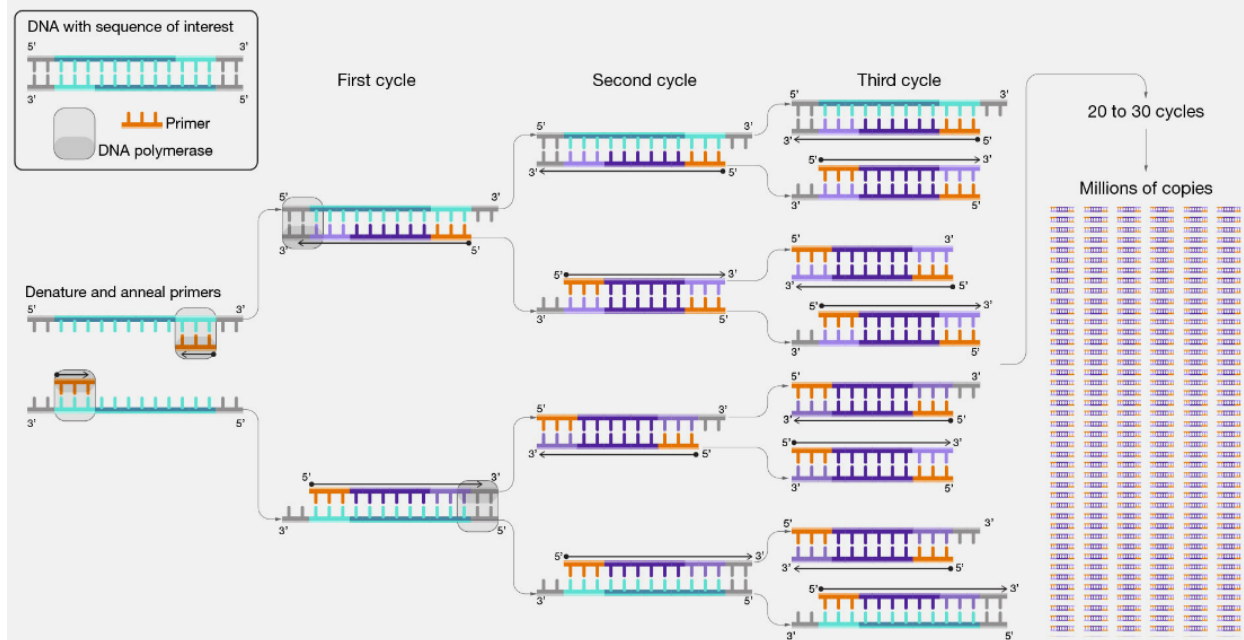
target DNA. PCR generally involves three steps: (1) denaturation of the double-stranded DNA, (2) annealing of primers, and (3) primer extension. Ex. 1030 at 1154.

54. During denaturation, double-stranded DNA molecules are separated into complementary single strands of DNA by using heat or other chemical or physical means of dissociation. *Id.* The two complementary strands of single-stranded DNA will be used as a template for replication.

55. During annealing, primers or probes, which are oligonucleotides complementary to a portion of one of the original DNA strands, attach to the single-stranded DNA templates. When the reaction mixture comprising denatured DNA template strands is cooled to a certain temperature, the primers hybridize, i.e. bind, through base pairing interactions to either the left (5') or right (3') side of the sequence of interest. *Id.* The portions of the templates to which the primers bind are called priming sites.

56. During primer extension, each molecule of single-stranded DNA is converted to double-stranded DNA, doubling the number of starting strands. A thermostable enzyme, called "DNA polymerase," catalyzes the synthesis of new strands of DNA through primer extension. The enzyme adds nucleotides complementary to those in the single-stranded DNA template starting at the 3' end of the annealed primer. *Id.* Upon completion of one cycle, the DNA strands are subjected to additional cycles, i.e., the same steps of denaturation, primer annealing,

and primer extension. This results in an exponential amplification of target DNA sequences, as exemplified below.



Ex. 1031 at 1.

57. Methods for applying PCR amplification to a small amount of sample polynucleotides were well established and known in the prior art. *See* Ex. 1014 at 5:37-42; Ex. 1008 at Abstract.

C. DNA Tagging

58. Tagging is a known concept in genetics that involves adding a synthetic polynucleotide to a sequence of interest. The '589 patent notes in its background section that DNA tagging was used in a variety of ways in the art. For example, the '589 patent refers to “a method of molecular barcoding to label template DNA prior to PCR amplification,” which “allows for the identification of contaminant and

redundant sequences by counting only distinctly tagged sequences.” Ex. 1001 at 2:12-16. The ’589 patent further acknowledges that various methods of tagging existed in the art. *See id.* at 2:20-22 (describing labeling by ligation); *id.* at 2:23-26 (describing labeling through specific primers); *id.* at 21:38-40 (“Tags may be associated with nucleic acid molecules by using, singly or in combination, ligation, nick-ligation, terminal transferase, hybridization, priming, and the like.”).

59. The field has been replete with routine methods of tagging for a variety of uses including as a unique “barcode” identifier (Ex. 1017 at 31:5-8); providing primers to facilitate amplification (Ex. 1008 at 762), counting and tracking molecules (Ex. 1014 at 9:23-24); detecting errors in amplification and sequencing (Ex. 1015 at 10:21-28); and/or removing amplification bias (Ex. 1007 at 1).

60. It was a common practice in the art to introduce molecular tags to label nucleic acids before amplification. *See, e.g.*, Ex. 1007 at 1; Ex. 1015 at 3:1-3; Ex. 1014 at 6:52-57; Ex. 1008 at 762. Indeed, the ’589 patent itself cites McCloskey as well as Miner et al., which neatly explains how tagging “allows for quantification of the relative abundance of genomic” patterns and sequences “by correcting for skewing that can arise from PCR amplification or the cloning of the products.” Ex. 1007 at 3.

61. Molecular tags have also been commonly known in the art to track the origin of nucleic acids. The concept of a “molecular barcode”—i.e., a barcode that

identifies the set of replication products derived from the same original molecule that was amplified, or a barcode that identifies each sample in a pool of samples, for example—was known in the prior art. This includes the use of nucleic acid sequences to track which patient a sample came from. *See, e.g.*, Ex. 1017 at 31:5-8. Porreca further discloses the use of a “differentiator tag sequence” which, when combined with target DNA templates, reduces the effects of bias by allowing for identification of unique input molecules after amplification. *Id.* at 28:49-67.

62. The concept of using multiple tags or a combination of tags was also known. For example, McCloskey discloses a method of DNA analysis for single-cell samples, among others, by utilizing an encoding oligonucleotide with five information regions, two of which are tag sequences – one being a barcode that uniquely identifies and distinguishes one polynucleotide from another and another being a batch-stamp which associates a polynucleotide to “a specific sample run on a specific date.” Ex. 1008 at 762, 764. Chee also discloses tagging a DNA sample with a combination of tags. Ex. 1015 at 20:12-15.

63. It was also known in the art that using a large number of tags is one way to ensure that each polynucleotide in a reaction mixture is tagged. For example, Chee discloses that the “large number of unique sequences that can be generated from the random assembly of nucleotides provides a high probability that each individual construct will be uniquely identified, even those constructs that contain

substantially the same” target DNA sequences. *Id.* at 11:26-30. Additionally, Porreca discloses using probes that “contain one of a large number of differentiator tags in their backbone such that the probability of any two [probes] in a reaction having the same differentiator tag sequence is low.” Ex. 1017 at 60:49-52. Brenner also discloses that “a size of a tag population is selected that ensures with high probability that substantially every target polynucleotide will have a unique tag.” Ex. 1014 at 18:2-5. Using a large number of tags was a typical and routine method of increasing the chances that every polynucleotide from the original sample will be tagged and eventually sequenced. McCloskey, for instance, gives a general description of how one can tune the length of the barcode to distinguish among a specific number of molecules. As McCloskey explains, the “number of distinguishable barcodes in a population of oligonucleotides used in a reaction is determined by the number of random bases, n . This enables one to distinguish among 4^n allele copies per reaction.” Ex. 1008 at 764.

64. Thus, tagging was a well-known step used as a prelude to sequencing of polynucleotides. In fact, Dr. Michael Wigler, named inventor of the '589 patent admitted in a scientific publication that “[m]odifying molecules to facilitate counting is not a new idea.” Ex. 1004 at 1. Dr. Wigler noted that the idea had been applied under a variety of names to identify PCR duplicates and improve counting of DNA

templates. *Id.* He further cited several references that disclose the modification of molecules to facilitate tagging, many of which are prior art to the '589 patent.

D. Amplification Bias and Molecular Tagging

65. One of the known issues with nucleic acid amplification is amplification bias. Because efficiency of the PCR process can vary for both systematic and stochastic—i.e., randomized—reasons, the relative quantity of each sequence in the post-amplification sample often may not accurately reflect the distribution of such samples in the pre-amplification sample. *See* Ex. 1032; Ex. 1033. For instance, if an original sample had two DNA molecules having a first DNA sequence and four DNA molecules having a second DNA sequence, this 1:2 ratio may not be preserved by amplification. PCR amplification may, for instance, favor the first DNA sequence over the second sequence (e.g., due to random failure of primer binding to the second sequence in early PCR cycles), such that after amplification one ends up with roughly 10,000 molecules of the first sequence and 10,000 molecules of the second sequence. Whereas the ratio was 1:2 prior to amplification, it is roughly 1:1 after amplification. Such problems cloud the ability of researchers or clinicians to understand, for example, how much of a particular gene is being expressed in a particular cell of interest or the copy number of a particular gene in the DNA of an input sample.

66. This amplification bias problem can be particularly severe when the amplification starts from a small number of template nucleic acid molecules. A low number of template DNA requires more rounds of amplification, which exponentially increases the effect of bias. Additionally, there is increased risk of amplifying any contaminant DNA in the starting mixture, which results in a skewed end product. Ex. 1007 at 1. Thus, reducing amplification bias is of particular importance when working with a limited DNA sample.

67. To solve this problem, it was well-known and very common to add molecular tags to target sequences prior to the amplification step. *See, e.g., id.* at 1; Ex. 1008 at 766; Ex. 1015 at 2:65-3:3; Ex. 1014 at 1:66-2:2; Ex. 1017 at 5:12-17. For example, one can use a large number of tags that each have a random DNA sequence, and the tags are randomly applied to the sample molecules. If there are multiple DNA molecules in the input sample that are the same, after the tagging process they will almost certainly be distinguishable from one another because the molecules will likely have all received a different tag assuming one uses a large enough number of tags. One can then amplify the tagged molecules and, if there is bias in the amplification, the bias can be corrected using the tags. This is because even if the amplification process preferentially amplifies particular DNA sequences over others, the total number of molecules of any given species that were present in the original sample can nonetheless be determined by counting the number of unique

tags associated with that sequence. This results in addressing amplification bias by keeping track of the polynucleotides in the original sample. Ex. 1007 at 1; Ex. 1008 at 766.

68. Tagging has been a known solution to address the problem of amplification bias. For example, McCloskey suggests use of a multiple tag system wherein sample polynucleotides derived from a sample are tagged so that after amplification “nonredundant” barcodes can be used to count the number of original molecules. *See* Ex. 1008 at 766 (“An additional benefit of our method for mosaic samples is that counting sequences with nonredundant barcodes provides a quantitative measure of template diversity.”).

69. The Miner reference teaches a similar process:

In response to the challenges of PCR redundancy and contamination associated with PCR amplification of limited amounts of DNA template, we have labeled genomic DNA fragments with molecular sequence barcodes and ‘batch-stamps’ prior to PCR amplification. This was accomplished by including these molecular labels in the hairpin linker sequence that we use in hairpin-bisulfite PCR (5). This encoded information enables us to track the genomic origin of each sequence obtained from PCR and subsequent bacterial cloning. ***Each genomic fragment is marked prior to amplification, allowing us to identify contaminant and redundant sequences and to quantify accurately the proportion of cells carrying a particular sequence variant by counting only distinctly tagged sequences.*** This highly sensitive method offers

confirmation of the independent genomic origin of all sequences in final data sets derived from PCR amplification.

Ex. 1007 at 1. As I explain further below, there are yet additional references that describe this tagging concept for counting.

E. Sequencing

70. Nucleic acid sequencing is the process of determining the sequence of nucleotides (A, T, C, and G) in a nucleic acid molecule. The development of sequencing technologies can be traced to at least 1977 when Frederick Sanger developed DNA sequencing technology based on a chain termination method. Ex. 1034 at 133, 140.

71. DNA sequencing technologies went through major improvements in subsequent years. In the early 2000s, next generation sequencing (“NGS”) technologies emerged that were different from the Sanger method in terms of throughput and cost. *Id.* at 133. By approximately 2008, techniques and platforms for NGS were commercially available and in widespread use including the Roche/454 FLX, Illumina/Solexa Genome Analyzer and Applied Biosystems SOLiD System. *Id.*

72. NGS is a type of high-throughput sequencing technology that uses “massively” parallel sequencing of many fragments of DNA. The platforms carrying out NGS are diverse in sequencing and detection biochemistry, but

massively parallel sequencing enables these sequencers to process and detect millions of sequence reads simultaneously. *Id.* DNA sequencing using NGS generally requires a minimum amount of DNA.

73. Once the sequences of such molecules have been obtained, they can be further analyzed using a so-called “reference genome.” *Id.* at 138. A reference genome is a genome sequence from the species that is being analyzed for which the sequence is already known. When sequence information from an unknown sample is determined, one can compare that unknown sequence information to the reference genome to determine where the unknown sequence came from in the species of the genome. The process of comparing an unknown sequence to a reference genome is known as “mapping” or “alignment.”

74. Molecular tags, whether used as barcodes, batch-stamps or for other purposes, are identified in the sequencing results. In McCloskey, for example, valid molecules were identified by detecting the valid batch-stamp sequence TCGAGCCT in the full molecular sequences. Ex. 1008 at 763. Similarly, contaminant molecules were identified through the invalid batch-stamp, GTACCAT. *Id.*

75. Although a substantial improvement over previous sequencing technologies, NGS sequencing is not perfect and can produce both random and systematic errors. As just one example, some sequencing techniques face detections

issues when sequencing long sequences of the same nucleotides (*e.g.*, distinguishing between TTTT and TTTTT). *See, e.g.*, Ex. 1034 at 133; Ex. 1017 at 57:10-14.

76. Sequencing errors were a known problem and one for which various solutions were developed. *See, e.g.*, Ex. 1015 at 10:54-57 (“In addition, error can be introduced as a result of sequencing error; this can be overcome by sequencing multiple instances of the same [target sequence]”); Ex. 1014 at 21:14-18 (“Preferably, sequences of oligonucleotide tags are selected so that they can be distinguished even in the presence of sequencing errors. Thus, oligonucleotide tags are selected to be maximally different.”).

F. Copy Number

77. Genes are sequences of DNA nucleotides and constitute functional units carrying hereditary information. Genes are located at specific positions on specific chromosomes. An individual typically inherits two copies of each gene, one from each parent.

78. In certain situations, however, copy number changes can occur such that genes are duplicated or deleted, resulting in a deviation from the standard two-copy state. Duplications of a gene could create more than two copies of a gene being present, while deletions could lead to one or zero copies of a gene being present. The concept of copy number simply pertains to the frequency with which a gene, or some other genetic region, is present in a person’s genome.

79. Copy number changes are important to the study of cancer, among other things. For example, a form of cancer might be caused by the deletion of a tumor suppressor gene in a cell. Various prior art techniques were directed towards identifying copy number changes, including the Brenner references that I rely upon here. Ex. 1014 at 1:55-60.

XI. THE PRIOR ART REFERENCES OF THE GROUNDS

A. Lo (Ex. 1010)

1. Lo is § 102 Prior Art

80. U.S. Patent Application Publication No. 2009/0029377 (“Lo”) entitled “Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing,” was published on January 29, 2009. Ex. 1010. I understand that Lo is prior art under pre-AIA 35 U.S.C. § 102(a) and (b) because Lo was published at least one year before the earliest claimed priority date of the ’589 patent. Lo is analogous art to the ’589 patent because it is in the same field of genetic analysis and is directed to the problem of using molecular counting to determine copy number information.

2. Lo Overview

81. Lo teaches a method for determining whether a chromosome or chromosome region imbalance exists within a biological sample. *Id.* ¶ 14. That is, Lo teaches a method for determining copy number information from a sample. Specifically, Lo discloses prenatal diagnostic testing methods for chromosomal

defects by “determining imbalances between different nucleic acid sequences”. *Id.* ¶ 3.

82. Fetal chromosomal aneuploidy is a condition where a fetus has, for instance, an abnormal number of chromosomes. *Id.* ¶ 4. The number of chromosomes “can be abnormally high, e.g. the presence of an extra chromosome 21 or chromosomal region in trisomy 21; or abnormally low, e.g. the absence of a copy of chromosome X in Turner syndrome.” *Id.*

83. Lo’s method can be applied to determine copy number information due to mutated genetic sequences, genetic polymorphisms, or copy number changes. *See, e.g., id.* ¶ 37 (“Yet other examples include sequences which are mutated, deleted, or amplified in a malignant tumor, e.g. sequences in which loss of heterozygosity or gene duplication occur.”); *id.* ¶109 (“In another example the sequencing technique on plasma cell-free DNA may be used to detect the chromosomal aberrations in the plasma DNA for the detection of a specific cancer. Different cancers have a set of typical chromosomal aberrations. Changes (amplifications and deletions) in multiple chromosomal regions may be used. Thus, there would be an increased proportion of sequences aligned to the amplified regions and a decreased proportion of sequences aligned to decreased regions.”).

84. Lo’s method begins with obtaining a biological sample, such as from a pregnant female or from a cancer patient. *Id.* ¶ 54. The sample contains nucleic acid

molecules from the fetus and the mother. *Id.* These nucleic acid molecules are fragments from chromosomes. *Id.* Lo discloses that this sample may be enriched or sub-selected for specific sequences before being sequenced in the next step. *Id.* ¶¶ 47, 72, 77-79. Next, the nucleic acid molecules from the biological sample are sequenced. *Id.* ¶ 55. The sequences of the molecules are frequently referred to in Lo as “sequenced tags” or just “tags.” To avoid confusion, it is important to note that Lo’s use of the term “tags” in this manner does not connote the “tags” that I refer to above, which are the short DNA molecules that are appended to DNA molecules for identification purposes. Rather, in Lo, the term “sequenced tags” or “tags” merely refers to a DNA sequence in the sample that comes from a particular chromosome.

85. Nevertheless, as Lo goes on to explain, the “resultant sequenced reads form a representative profile of the mix of nucleic acid species in the original specimen.” *Id.* ¶ 57. Lo teaches that an imbalance in the number of sequence reads coming from a chromosome or chromosome region is indicative of a copy number change and can be inferred from the percentage representation of the sequences from the region of interest relative to the other “sequenced tags” of the specimen. *Id.* ¶ 67.

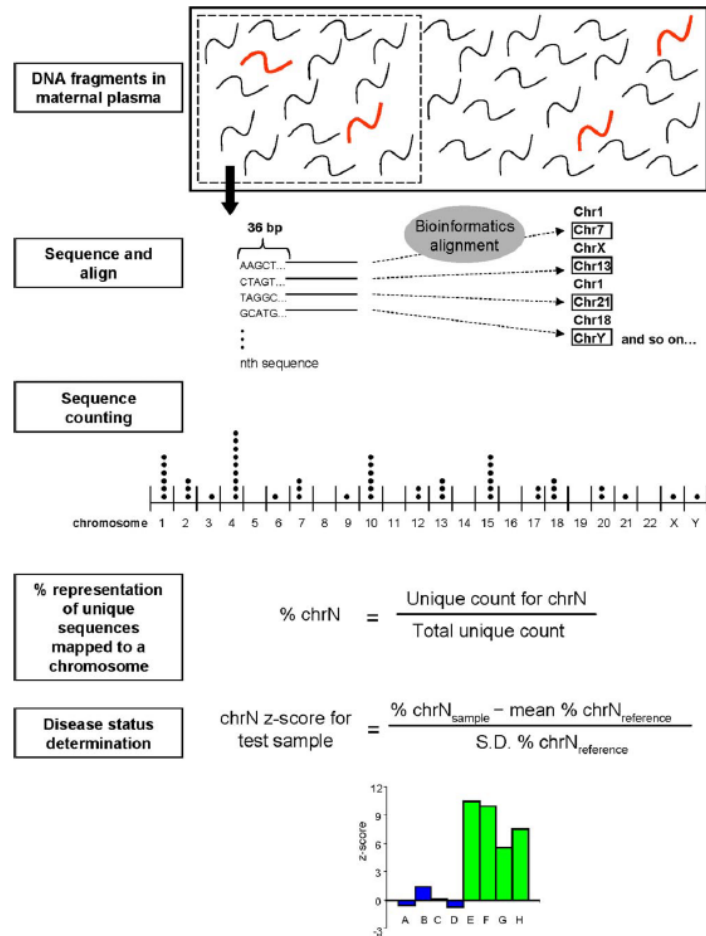
86. Lo discloses several ways to assess the percentage representation of a chromosome or chromosome region. *Id.* ¶ 60. These can include “counting the number of sequenced tags, the number of sequenced nucleotides (basepairs) or the

accumulated lengths of sequenced nucleotides (basepairs) originating from particular chromosome(s) or chromosomal regions.” *Id.* In one embodiment, Lo teaches aligning the sequenced tags to a human reference genome sequence to determine chromosomal origin. *Id.* ¶ 70. This process of “aligning” is nothing more than the process of determining which particular chromosome or chromosome region a DNA sequence came from, and long before the priority date of the ’589 patent.

87. From there, the total number of individual “sequenced tags” aligned to each chromosome or chromosome region can be tabulated. *Id.* Lo then teaches determining a ratio of the number of sequence reads from the chromosome or chromosome region of interest to some other “reference” chromosome or group of chromosomes within the same sample that one expects will not have a copy number change. *Id.* ¶ 62. As Lo explains, this ratio may be a percentage of the total number of reads coming from the chromosome or chromosome region of interest relative to all of the other reads. *Id.* ¶ 63. Finally, this ratio or percentage is compared to one or more cutoff values, which in turn, allows for a determination of whether a copy number change exists for the first chromosome or chromosomal region. *Id.* ¶¶ 64, 66. Lo discloses that the cutoff values are determined using an external standard based on normal genomes. As Lo states, the “parameter (*e.g.*, the fractional representation of the clinically relevant chromosome) is then compared to a

reference range established in pregnancies involving normal (i.e. euploid) fetuses.” *Id.* ¶ 65; *see also id.* ¶ 70 (“The total number of individual sequenced tags aligned to each chromosome were tabulated and compared with the relative size of each chromosome as expected from the reference human genome or non-disease representative specimens. Chromosome gains or losses were then identified.”); *id.* ¶ 95 (“the chromosome 21 over-representation could be detected by comparing the percentage representation of chromosome 21 obtained experimentally with the percentage representation chromosome 21 sequences expected for a euploid human genome.”).

88. FIG. 1 of a related 2008 publication by Lo (Ex. 1011) neatly explains the method in Lo:



Id. Fragments of fetal DNA are highlighted in red. From each fragment, 36 basepairs are sequenced. That short sequence is used to identify which chromosome or chromosomal region the fragment comes from. The count of the number of fragments from each chromosome or region is compared with the total count of fragments to compute a percentage of sequences mapped to a particular chromosome or chromosomal region. That calculated percentage is compared against the percentage for that chromosome or chromosome region in a reference standard genome. Samples containing several standard deviations more chromosomal

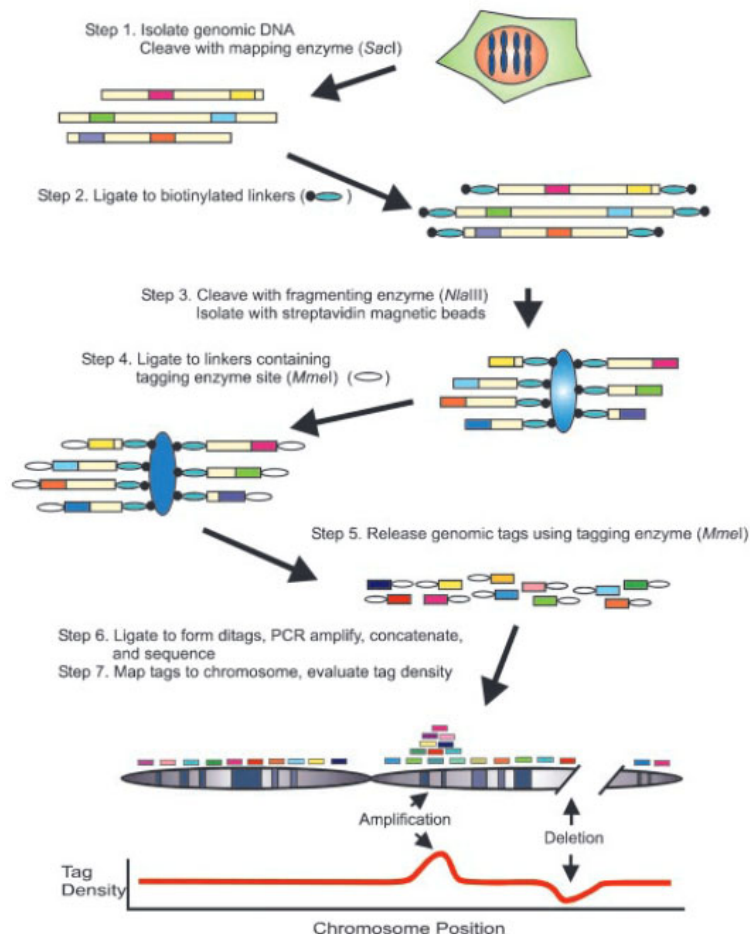
content than the reference genome are likely aneuploidy and are highlighted in green.

89. The method set forth in Lo based on looking for an imbalance in the number of sequence reads that map to a particular chromosome or chromosome region was not by any means a new method, but was rather an old idea that had been applied in many other contexts. Lo, in fact, acknowledges that the prior art discloses other such methods.

90. Lo, for instance, explains that the prior art Dhallan references teach similar methods: According to Lo, in Dhallan's method, the "proportion of chromosome 21 sequences contributed by the fetus in maternal plasma was determined by assessing the ratio of paternally-inherited fetal-specific alleles to non-fetal-specific alleles for single nucleotide polymorphisms (SNPs) on chromosome 21. SNP ratios were similarly computed for a reference chromosome. An imbalance of fetal chromosome 21 was then inferred by detecting a statistically significant difference between the SNP ratios for chromosome 21 and those of the reference chromosome...." Ex. 1010 ¶ 8.

91. While the examples and embodiment in Lo focuses on the use of cell-free DNA from pregnant women, wherein the sample comprises DNA from multiple genomes (i.e. mother and fetus), the basic counting method described in Lo was a standard method that had been previously applied straightforwardly to other sources

of DNA to detect copy number information. For example, Wang et al. (Ex. 1012) describes a tagging and sequencing method to detect copy number changes called “digital karyotyping.” Wang and co-workers described characterizing “210,245 genomic tags from the lymphoblastoid cells of a normal individual (NLB) and 171,795 genomic tags from the colorectal cancer cell line (DiFi).” *Id.* at 16158. In step 7 of FIG. 1, reproduced below, an abnormal number of sequences mapped to a chromosomal region indicates amplification while an absence indicates deletion:



Id. at 16157. The bottom of the figure with the chart labeled “Tag Density” most clearly illustrates how Wang’s method identifies copy number information (either an amplification or a deletion) by looking for imbalances in the number of sequence reads mapping to a region of a chromosome.

92. As another example, a 2009 paper by Chiang et al. (Ex. 1013) titled “High-resolution mapping of copy-number alterations with massively parallel sequencing” describes a similar segmentation algorithm to identify copy number changes using a variable window size. *Id.* at 100-101. This method was applied to sequencing data “from three tumor cell lines...and their matched normal cell lines.” *Id.* at 101. Sequencing counts in a given genomic window were compared between the tumor and reference samples to identify “194 copy-number alterations in the HCC1954 cell line, 126 alterations in the HCC1143 cell line and 15 alterations in the NCI-H2347 cell line.” *Id.*

93. Thus, while the Lo publication focuses primarily on applications where the sample was cell-free DNA, I wish to make clear that there is nothing about the method in Lo that would limit it to such applications. In fact, a POSA would immediately recognize and understand that Lo’s method is applicable to other sources of DNA and had, in fact, been applied to other DNA sample types that were not cell-free DNA well before Lo’s patent publication.

B. McCloskey (Ex. 1008)

1. McCloskey is § 102 Prior Art

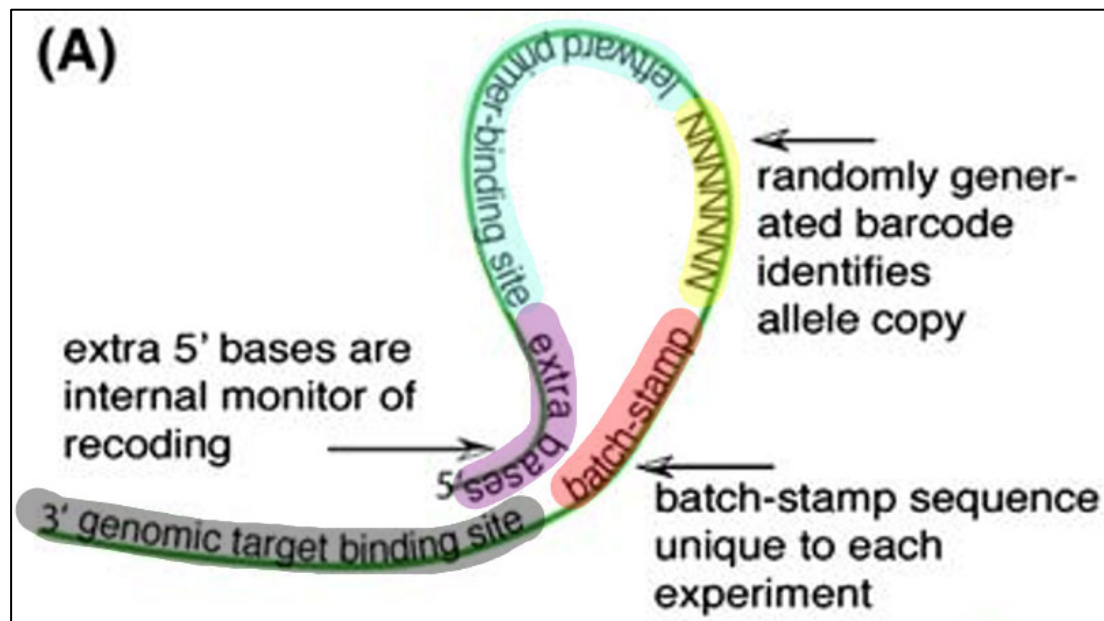
94. McCloskey *et. al.*, “Encoding PCR Products with Batch-Stamps and Barcodes,” is an article that was published in Biochemical Genetics on October 23, 2007. Ex. 1008. I understand that McCloskey is thus prior art to the ’589 patent under pre-AIA 35 U.S.C. § 102(a) and 102(b).

2. McCloskey Overview

95. McCloskey is one of four references that I rely upon in this declaration that teaches the general approach of tagging, amplification, and sequencing for molecular counting. McCloskey is analogous art to the ’589 patent because, as documented herein, it is in the field of genetic analysis using sequencing and addresses the problem of molecular counting using sequencing.

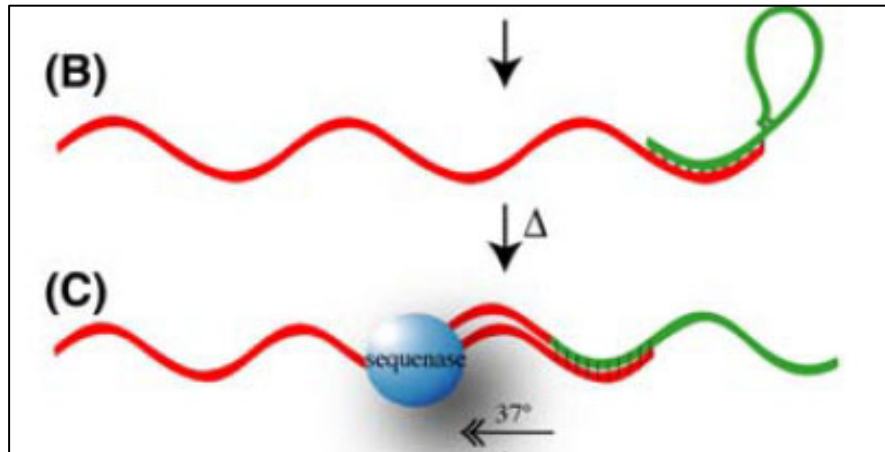
96. McCloskey describes a method of DNA analysis for single-cell samples (among other sample types) that utilizes an “encoding oligonucleotide” having “five distinct information regions,” two of which are tag sequences that distinguish, respectively, polynucleotides and samples. *Id.* at 762. As depicted below, the first tag is a “barcode” (highlighted in yellow), which is a randomly generated sequence intended to uniquely identify and distinguish one polynucleotide from another. *Id.* (The “random barcode distinguishes among sequences arising from different cell or allele copies.”). McCloskey also teaches a second tag sequence, the “batch-stamp” (highlighted in red), that is unique to each experiment and is intended to associate a

polynucleotide to “a specific sample run on a specific date.” *Id.* at 764. The other regions on the encoding oligonucleotide include: “a genomic target-binding site” (highlighted in grey), “a primer-binding site for PCR” (highlighted in light blue), and “extra bases that serve as internal sentinels of recoding as well as tethering the oligonucleotide to itself” to minimize non-specific annealing (highlighted in purple, below). *Id.* at 763.



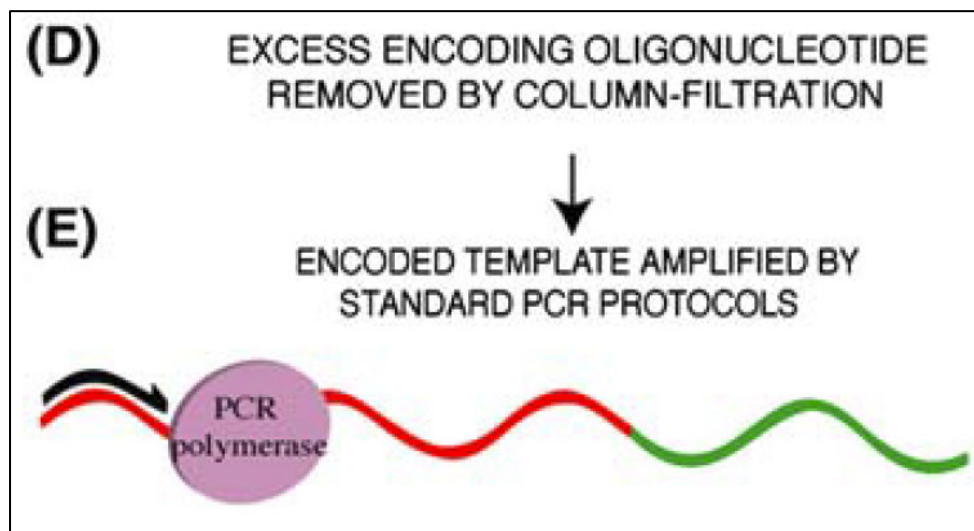
Id. at FIG. 1.

97. McCloskey’s method begins with annealing the 3’ end genomic target binding site (highlighted in grey, above) of the encoding oligonucleotide to denatured genomic DNA (see step (B) in figure, below). *Id.* at 763-766. The oligonucleotide-annealed DNA is then extended to build a complementary DNA sequence (see step (C) in figure, below):



Id. at FIG. 1.

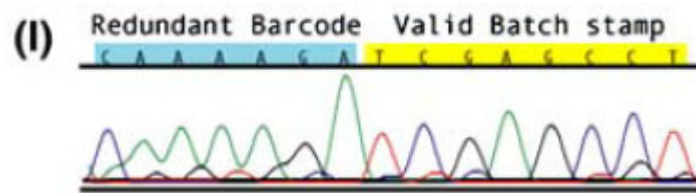
98. The resulting sample of encoded DNA strands are then amplified using PCR and sequenced for evaluation (see step (E) in figure, below). *Id.* at 765-766.



Id. at FIG. 1.

99. McCloskey's method addresses the element of uncertainty in sequence information caused by the risks of amplification bias, contamination, and redundant sequences. *Id.* at 762. McCloskey teaches that a benefit of its method is that it distinguishes redundant sequences found in the end product of PCR amplification

from redundant sequences existing in the sample before amplification. Researchers can readily identify the redundant sequences from the PCR reaction “because they are identical in both barcode and batch-stamp” as opposed to any encoded nucleotides with different barcodes, which suggests that the redundancy existed in the sample prior to amplification (see step (I) in figure, below). *See id.* at 766. Thus, by “counting sequences with non-redundant barcodes”, McCloskey’s method is useful in obtaining “a quantitative measure of template diversity.” *Id.*



Id. at FIG. 1. The McCloskey method based on the use of the random barcodes is thus little different from the claimed method in the ’589 patent.

C. Brenner (Ex. 1014)

1. Brenner is § 102 Prior Art

100. U.S. Patent No. 7,537,897 (“Brenner”) entitled “Molecular Counting” was published on May 26, 2009. Ex. 1014. I understand that Brenner is thus prior art to the ’589 patent under pre-AIA 35 U.S.C. § 102(a) and (b) because Brenner was published at least one year before the earliest claimed priority date of the ’589 patent.

2. Brenner Overview

101. Brenner is the second of four references that I rely upon in this declaration that teaches the general approach of tagging, amplification, and sequencing for molecular counting. Brenner is analogous art to the '589 patent because, as documented herein, it is in the field of genetic analysis using sequencing and addresses the problem of molecular counting using sequencing

102. Brenner teaches methods for counting molecules in a sample using unique oligonucleotide tags. *Id.* at Abstract. Brenner's purported approach is to convert the problem of counting molecules into the problem of counting tags. *Id.*; *see also id.* at 1:64-66 ("The invention provides a method of counting molecules in a sample by converting the problem of counting molecules into one of counting sequences of oligonucleotide tags."). Brenner further discloses various methods of counting these tags, including directly sequencing a sample of tags. *Id.* at Abstract; 2:16-20 ("In one aspect, the different oligonucleotide tags of the selected probes, and hence, the number of target molecules, is determined by sequencing a sample of the oligonucleotide tags amplified from the selected probes.").

103. Brenner's method purports to address the need for cost-effective, convenient, and accurate digital measurements of quantities of biomolecules. *Id.* at 52-55. Brenner discloses how this method has applications "in medical and research fields for determining a wide variety of quantities, including genetic copy number variation." *Id.* at 1:56-58; *see also id.* at 4:4-9 ("The present invention provides

compositions and methods for making digital measurements of biomolecules, and has applications in the measurement of genetic copy number variation...”). Brenner’s method employs conventional techniques of organic chemistry, polymer technology, molecular biology, cell biology, biochemistry, and immunology, known to those skilled in the art at the time of the invention. *Id.* at 4:58-63. These conventional techniques included polymer array synthesis, hybridization, ligation, and detection of hybridization using a label. *Id.* at 4:63-66.

104. Brenner teaches a method for estimating the number of target polynucleotides in a mixture. *Id.* at 3:15-24. First, target polynucleotides are labeled with unique oligonucleotide tags. *Id.* The target polynucleotides “are each associated or linked, e.g. by ligation, to an oligonucleotide tag from a population that has a much larger size than that of the target polynucleotide population.” *Id.* at 17:61-64. “Generally, a size of tag population is selected that ensures with high probability that substantially every target polynucleotide will have a unique tag.” *Id.* at 18:2-4.

105. To these ends, Brenner further discloses methods for constructing sets of tags and attaching tags to polynucleotides. *Id.* at 12:55-56 (“A general procedure for attaching oligonucleotide tags to polynucleotides is illustrated in FIGS. 2A-2B.”); *id.* at 24:51-25:67 (“Nucleotide sequences of sets of such oligonucleotide tags are readily generated by conventional computer programs that generate all possible

sequences of a preselected length followed by sorting all prospective tag sequences that fulfill the desire criterion on homopolymer size.”); *id.* at 28:35-30:4 (Example I: Construction of a Metric Tag Set); 30:6-31:40 (Example II: Construction of a Binary Tag Set).

106. After the tagged polynucleotides are generated, the tags are amplified using PCR. *Id.* at 3:20-21 (“amplifying the oligonucleotide tags of the labeled target polynucleotides”); *id.* at 17:52-55 (“In this latter approach, oligonucleotide tags may be amplified using emulsion PCR, e.g. as also disclosed in the cited references, so that clonal populations of each oligonucleotide tag of a sample are formed on beads, which are then sequenced.”); *id.* at 23:36-38 (“Once such an emulsion is formed, conditions are selected for implementing an amplification reaction, such as PCR....”).

107. Finally, the number of different oligonucleotide tags in the sample of amplified oligonucleotide tags are counted to determine the number of target polynucleotides in the starting sample. *Id.* at 2:2-4 (“The number of different oligonucleotide tags detected, or counted, is equal to the number of molecules in the sample.”); *id.* at 3:30-34 (“[T]he number of different oligonucleotide tags in a sample is determined by determining the nucleotide sequences thereof and then counting the number of oligonucleotide tags with different sequences.”); *id.* at 5:17-23 (“[S]ubstantially every molecule to be counted in a sample, e.g. the number of

single stranded DNA molecules of a particular genetic locus in a sample of genomic DNA, is associated with a probe having a different tag, so that the process of counting multiple copies of the same molecule is transformed into a process of counting the number of different kinds of associated tags.”).

108. In particular, Brenner teaches methods of counting by directly sequencing samples of oligonucleotide tags. *Id.* at Abstract (“[D]ifferent tags may be counted by directly sequencing a sample of tags using any one of several different sequencing methodologies.”); *id.* at 22:19-22 (“The oligonucleotide tags are then amplified **(5252)** out of conjugates **(5246)**, after which a sample of such tags are sequenced and the number of different tags are determined **(5254)**.”); *id.* at 24:30-35 (“The size of the sample of oligonucleotide tags from selected probes that is sequenced is a design choice in that the more oligonucleotide tags that are sequenced the greater the probability that every different oligonucleotide tag (and hence the correct number of target polynucleotides) will be determined.”).

D. Chee (Ex. 1015)

1. Chee is § 102(e) Prior Art

109. U.S. Patent No. 9,085,798 (“Chee”) entitled “Nucleic Acid Constructs and Methods of Use” was published on July 21, 2015. Ex. 1015. Chee issued from U.S. Patent Application No. 13/266,568, which is the national phase entry of International PCT Application No. US2010/033064 (“the Chee PCT”), filed on April

29, 2010. Chee and the Chee PCT share a specification. The Chee PCT designated the United States and was published in English. Ex. 1016. As such, I understand from counsel that Chee is prior art to the '589 patent under pre-AIA 35 U.S.C. § 102(e).

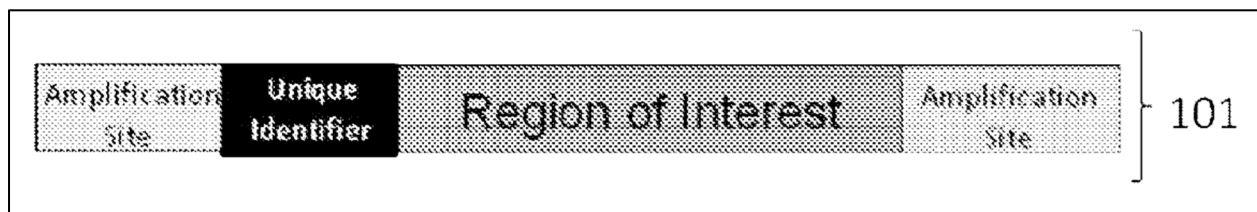
2. Chee Overview

110. Chee is the third of four references that I rely upon in this declaration that teaches the general approach of tagging, amplification, and sequencing for molecular counting. Chee is analogous art to the '589 patent because, as documented herein, it is in the field of genetic analysis using sequencing and addresses the problem of molecular counting using sequencing

111. Chee teaches counting uniquely tagged molecules for “identifying potential sample bias introduced by quantitative analysis techniques.” Ex. 1015 at 5:19-20. Chee’s purported approach relies on so-called “oligonucleotide constructs.” *Id.* at Abstract (“The present invention provides oligonucleotide constructs, sets of such oligonucleotide constructs, and methods of using such oligonucleotide constructs . . .”). These nucleic acid constructs comprise a unique identifier (or “tag”), a region of interest (“ROI”), and at least one amplification site. *Id.* at 2:25-33. Chee teaches creating these constructs by tagging nucleic acids derived from a biological sample. *Id.* at 12:17-21 (“[W]hen the constructs are to be used in quantitative analysis of nucleic acids derived from a biological sample, the

nucleic acids containing the ROI are isolated, and the additional elements of the constructs are separately created and added to the sample nucleic acids using molecular biology techniques.”); *see also id.* at 4:1-3 (“In another aspect the constructs are created using fragmented nucleic acids, and the unique identifier and amplification regions are added to the individual fragments”); *id.* at 11:18-20 (“[A] diverse but pre-defined set of sequences can be synthesized and attached to the ROIs, e.g., via PCR primers or by ligation.”).

112. An example of a Chee construct can be seen in the below excerpt of Figure 1:



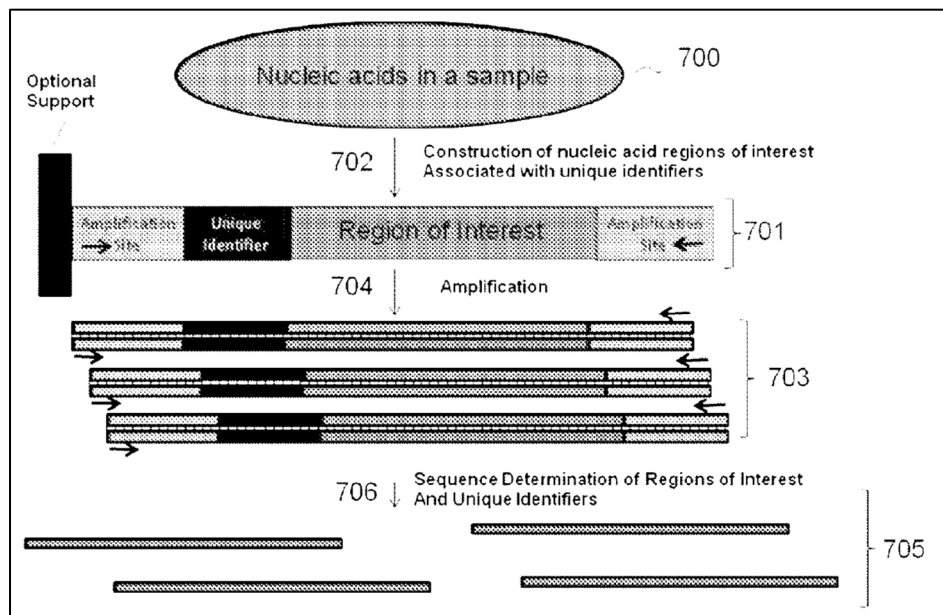
Id. at FIG 1.

113. Chee further discloses amplifying the set of constructs. *Id.* at 13:21-25 (“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.”); *id.* at 13:33-37 (“In one example, universal primers that are complementary to the amplification sites in the constructs are used to perform a

limited number of amplification cycles, so that each unique molecule now has a small number of identical copies each containing an ROI and the unique identifier.”).

114. Chee further discloses that this amplification can be carried out using conventional PCR methods. *Id.* at 13:40-43 (“Amplification can be carried out in solution, using techniques such as PCR or emulsion PCR (with or without beads), and/or using solid-phase amplification, e.g., on a support surface.”); *id.* at 14:33-36 (“In one particular aspect, the individual molecules are cloned onto beads and amplified using emulsion PCR to produce many copies of the original template on the surface of a bead.”); *see also id.* at 3:13-17 (“These constructs allow the unique identifier and the ROI to be selectively amplified using single primer-dependent amplification techniques such as asymmetric PCR.”).

115. Following amplification, Chee discloses using sequencing methods to confirm the sequence of the ROI and the unique identifier of a particular construct. *Id.* at 14:3-5 (“[C]omplete constructs are sequenced to confirm both the ROI and the unique identifier of a particular construct.”). The below excerpt from Figure 7 shows tagging, amplification, and sequencing as taught by Chee.



Id. at FIG. 7. Chee teaches that “molecules with identical identifiers arise via amplification of the same original construct.” *Id.* at 14:44-46.

116. Chee further discloses counting the number of constructs with unique identifiers. *Id.* at 9:13-21 (“The ability to uniquely identify specific ROI-containing constructs (as opposed to identifying the same ROI present in multiple constructs) . . . allows the identification of technical bias in, say, amplification procedures used in quantitative analysis, *as a skewed proportion of a specific construct in a sample can indicate a technical bias.*”); *id.* at 10:41-46 (“[T]he unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be *inferred by association of each initial ROI present in a sample with its unique identifier.*”). As Chee explains, “[b]y creating constructs comprising the nucleic acids directly isolated from a sample, bias due to subsequent manipulation

of the molecules,” e.g., bias in amplification of certain sequences, “can be identified by determining the amounts of constructs with unique identifiers.” *Id.* at 23:56-59. “Thus, if a nucleic acid is either over-represented or under-represented due to amplification bias, the number of molecules comprising unique identifiers will also be over-represented or under-represented, and the quantification of the molecules can be corrected to reflect this manner of bias.” *Id.* at 23:59-64.

E. Porreca (Ex. 1017)

1. Porreca is § 102(e) Prior Art

117. U.S. Patent No. 11,840,730 (“Porreca”) entitled “Methods and Compositions for Evaluating Genetic Markers” was published on December 12, 2023. Ex. 1017. I understand that Porreca claims priority to at least International PCT Application No. US2010/001293 (“the Porreca PCT”), filed on April 30, 2010. *Id.* at (63); *see* Ex. 1018. Porreca and the Porreca PCT share a specification. The Porreca PCT designated the United States and was published in English. *Id.* I understand from counsel that Porreca is therefore prior art to the ’589 patent under pre-AIA 35 U.S.C. § 102(e).

2. Porreca Overview

118. Porreca is the fourth of four references that I rely upon in this declaration that teaches the general approach of tagging, amplification, and sequencing for molecular counting. Porreca is analogous art to the ’589 patent

because, as documented herein, it is in the field of genetic analysis using sequencing and addresses the problem of molecular counting using sequencing.

119. Porreca teaches counting uniquely tagged nucleic acid molecules for determining genotypes in patient samples, with a focus on reducing bias and increasing the reproducibility of multiplex genetic analysis. Ex. 1017 at Abstract (“Aspects of the invention relates to methods and compositions that are useful to reduce bias and increase the reproducibility of multiplex analysis of genetic loci.”); *id.* at 1:23-24 (“The invention relates to methods and compositions for determining genotypes in patient samples.”).

120. Porreca’s purported approach uses nucleic acid tags (referred to as differentiator tags) to tag target nucleic acid molecules. *Id.* at 5:1-5 (“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.”). These differentiator tags allow for “the detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reactions.” *Id.* at 5:5-8.

121. Porreca’s purported method is useful for “detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats.” *Id.* at 55:50-51; *see also id.* at 20:21-25 (“Methods of the invention may be used for any type of mutation, for example a single base change (e.g., insertion, deletion,

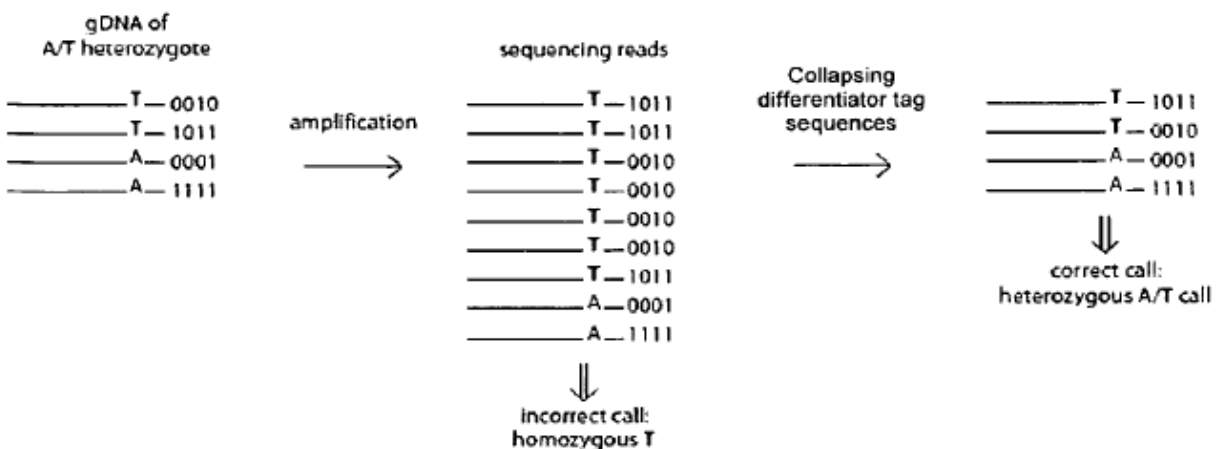
transversion or transition, etc.), a multiple base insertion, deletion, duplication, inversion, and/or any other change or combination thereof.”).

122. Porreca discloses tagging target nucleic acid sequences of interest with differentiator tags. *Id.* at 29:4-6 (“FIG. 4A-C depict various non-limiting examples of methods for combining differentiator tag sequence and target sequences.”); *id.* at 29:28-33 (“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).”); *id.* at 31:10-13 (“The actual incorporation of the random differentiator tag sequences can be performed through various methods known in the art. For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation.”).

123. Porreca further discloses the use of amplification and sequencing techniques on the tagged molecules. *See, e.g., id.* at 72:11-12 (“Amplify Circular Material by PCR Using Primers Specific to the ‘Constant Region’ of the Probes”); *id.* at 72:45-47 (“Perform Sequencing of Library According to Manufacturer's Directions (e.g. Illumina, ABI, Etc), Reading Both the Target Sequence and the Differentiator Tag Sequence.”); *see generally* 71:24-73:5 (Example 2).

124. Porreca further discloses counting differentiator tags by sequencing. *Id.* at 60:65-67 (“The number of each differentiator tag sequence is counted, for example, by next-generation sequencing.”); *id.* at 72:50-51 (“Construct a table of target: differentiator tag abundances from the read data....”). The differentiator tags are counted “to distinguish independent target molecules from each other....” *Id.* at 28:29-30. “[I]f 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.” *Id.* at 28:37-42.

125. Figure 5, reproduced below, summarizes the foregoing method of counting by the using of tagging, amplification, and sequencing:



Id. at FIG. 5; *see also id.* at 72:63-73:2 (“All ‘count’ entries should be ‘1’, since 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. For any target:differentiator tag combination observed more than once, all such reads are ‘collapsed’ into a single read before consensus basecalls are determined.”).

F. Quake (Ex. 1019)

1. Quake is Prior Art

126. U.S. Patent No. 8,195,415 (“Quake”) entitled “Noninvasive Diagnosis of Fetal Aneuploidy by Sequencing” issued on June 5, 2012 from U.S. Patent Application No. 12,696,509, filed on January 29, 2010 and published on June 3, 2010. Ex. 1019. I understand from counsel that Quake is thus prior art to the ’589 patent under pre-AIA 35 U.S.C. § 102(e).

2. Quake Overview

127. Quake teaches a method for the noninvasive diagnosis of fetal aneuploidy by DNA sequencing that shares many similarities with Lo’s method, i.e. it involves sequencing mixed-genome samples (i.e. from mother/fetus) in unknown ratios. *Id.* Just like Lo’s method, in Quake’s method, DNA is first obtained from a maternal blood sample. *Id.* at 3:43-45. The DNA is partially sequenced to create a large number of short reads, which Quake refers to as “sequence tags.” *Id.* at 3:45-50. These tags are sufficiently unique to be mapped to specific chromosomes or chromosomal locations known to exist in the genome. *Id.* The tags are mapped to

their corresponding location on the chromosome using a reference sequence. *Id.* at 4:40-43.

128. The Quake method is likewise similar to the Lo method in that it is based on determining an imbalance in the number of sequence reads in a region of interest in the genome. As Quake states, fetal aneuploidy can be detected by “counting the number of sequence tags mapped to each chromosome” and noting “the over- or under-representation of any chromosome or chromosome portion.” *Id.* at 3:50-56.

129. The difference between Quake and Lo has to do with the standard that is used to assess whether a sample has a copy number change. Lo’s method relies upon an external standard derived from samples that do not have a copy number change. Quake’s method, in contrast, does not require the use of an external standard and instead uses an internal standard. *See id.* at 9:6-10 (“This method does not require the use of an external standard, but, rather, provides an internal reference, derived from al [sic] of the sequence tags (genomic sequences), which may be, for example, a single chromosome or a calculated value from all autosomes.”).

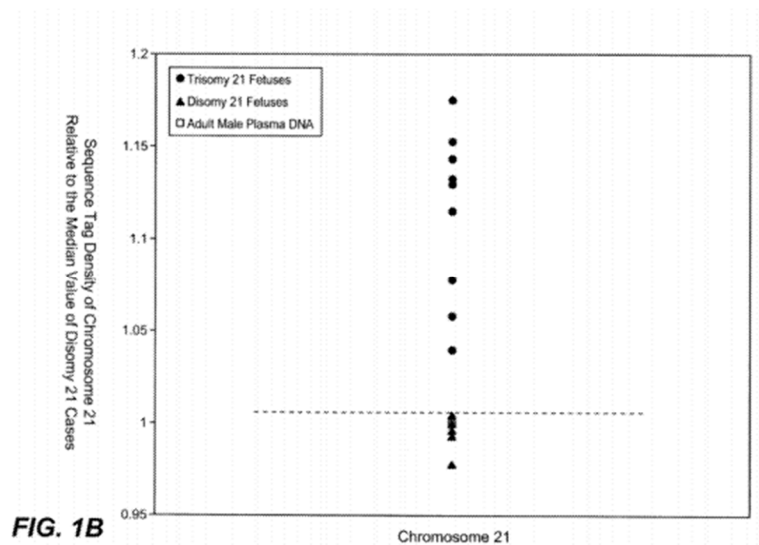
130. More specifically, Quake divides the genome into “windows” or “bins.” For example, many of the examples in Quake use bins that are 50 kb in length, but Quake contemplates other sizes. Then, one computes the median number of sequence reads mapping to all the bins across the chromosomes. As Quake

explains, by “taking the median M of all 22 autosomes, one obtains a normalization constant N that can be used to correct for differences in sequences obtained in different runs, as can be seen in Table 1. Thus, the normalized sequence tag density for chromosome 1 would be $M1/N$; for chromosome 22 it would be $M22/N$.” *Id.* at 7:44-54.

131. Similarly, Quake explains, the “value of the sequence tag density is normalized within a sample. This can be done by counting the number of tags falling within each window on a chromosome; obtaining a median value of the total sequence tag count for each chromosome; obtaining a median value of all of the autosomal values; and using this value as a normalization constant to account for the differences in total number of sequence tags obtained for different samples. *A sequence tag density as calculated in this way would ideally be about 1 for a disomic chromosome.*” *Id.* at 8:61-9:4.

132. Thus, rather than using an external control, Quake looks at many regions across the genome in an individual sample, most of which will not include a copy number change, and then computes normalized densities. A region of the genome with density that deviates from the normal sequence tag density of “about 1” can be identified as having a different copy number. This is shown, for instance, in Figure 1B, which shows that in samples where there is an extra copy of

chromosome 21 the normalized sequence tag density is substantially above the ideal normal of 1:



Id. at FIG. 1; *see also id.* at Abstract (“The median count of autosomal values is used as a normalization constant to account for differences in total number of sequence tags is used for comparison between samples and between chromosomes.”).

133. Quake further teaches that its method is not limited to entire chromosomes. Specifically, Quake teaches that to increase sensitivity its method can focus on certain regions of the genome by selecting in advance DNA sequences from regions of interest. As Quake states, “[a]nother method for increasing sensitivity to fetal DNA is to focus on certain regions within the human genome. One may use sequencing methods which select a priori sequences which map to the chromosomes of interest (as described here elsewhere, such as 18, 21, 13, X and Y). One may also choose to focus, using this method, on partial chromosomal deletions,

such as 22q11 deletion syndrome. Other microdeletions and microduplications are set forth in Table 1 of US 2005/0181410, published Aug. 18, 2005 under the title ‘Methods and apparatuses for achieving precision genetic diagnosis.’” *Id.* at 13:53-63.

XII. PRIOR ART RENDERS OBVIOUS THE CLAIMS OF THE ’589 PATENT

134. I now discuss the specifics of how the prior art renders obvious the claims of the ’589 patent.

135. The basic structure of my theory is as follows. Lo teaches the same general approach set forth in claim 1 of the ’589 patent, wherein one determines copy number information by sequencing genomic segments, mapping the sequence reads to a reference genome, and then finally determining copy number information by counting sequence reads for the genomic region of interest.

136. To the extent any aspect of the independent claims is missing in Lo, it is the process of tagging the DNA sequences to generate unique tagged nucleic acid molecules followed by amplification of these tagged nucleic acid molecules. Yet, as I explained above and further below, this process was well-known to the skilled artisan and was a basic technique in their toolkit to improve molecular counting. Skilled artisans would have been motivated to use this basic tagging and amplification technique to improve the accuracy of molecular counting, which is what Lo’s method is based upon. Indeed, using tagging and amplification with Lo

would have been the simple application of a known method to improve an existing process, and a POSA would have had a reasonable expectation of successfully doing so given the wide and routine use of such methods. Four prior art references expressly instruct to apply the tagging and amplification method to determine copy number information by molecular counting. As such, claim 1 of the '589 patent is obvious in view of Lo and other prior art references.

137. As to the dependent claims in the '589 patent, they recite features that are already part of Lo's method or that are otherwise similarly well-known obvious techniques that a skilled artisan would have been motivated to use with a reasonable expectation of success.

A. Ground 1: Claims 1-3, 6-8, 11, and 16-18 Are Unpatentable as Obvious Over Lo in View Of The Knowledge of a POSA As Reflected In McCloskey, Brenner, Chee, And Porreca

1. Claim 1:

a) [1A] A method comprising: a) obtaining segments of genomic nucleic acids from a sample;

138. It is my opinion that Lo discloses a method comprising the step of "obtaining segments of genomic nucleic acids from a sample."

139. First, Lo discloses receiving a biological sample, such as from a pregnant female, and states the "biological sample may be plasma, urine, serum, or any other suitable sample. The sample contains nucleic acid molecules from the

fetus and the pregnant female. For example, the nucleic acid molecules may be fragments from chromosomes.” Ex. 1010 ¶ 54.

140. Further, Lo states that “the term ‘biological sample’ as used herein refers to ‘any sample that is taken from a subject (e.g., a human, such as a pregnant woman) and contains one or more nucleic acid molecule(s) of interest.’” *Id.* ¶ 33. Lo additionally teaches obtaining maternal plasma samples and that such samples include DNA that is known to be naturally fragmented, *e.g.*, “[i]n one embodiment, random sequencing is performed on DNA fragments that are present in the plasma of a pregnant woman, and one obtains genomic sequences which would originally have come from either the fetus or the mother.” *Id.* ¶ 83; *see also id.* ¶ 55 (“In step 120, at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced...In one embodiment, the nucleic acid molecules are fragments of respective chromosomes. One end ... both ends, or the entire fragment may be sequenced.”).

141. Therefore, in my opinion Lo teaches that the first step in the method of obtaining a sample containing nucleic acid molecules or segments.

- b) [1B] tagging the segments with nucleic acid tags, thereby generating a plurality of unique tagged nucleic acid molecules, such that each of the unique tagged nucleic acid molecules comprises a segment of the segments of the genomic nucleic acids from step (a) and a nucleic acid tag of the nucleic acid tags;**

142. It is my opinion that it would have been obvious to perform this claim step in combination with the method set forth in Lo. This would have been an obvious combination in view of the knowledge of the skilled artisan, as reflected in McCloskey, Porreca, Chee, or Brenner.

143. As of the earliest priority date of the '589 Patent, it was well-known and routine in the art to attach tag sequences to nucleic acid fragments prior to amplification to obtain accurate genomic copy number information. Indeed, the specification of the '589 Patent admits as much in its discussion of Miner, McCloskey, and Brenner:

Parameswaran et al. (2007) and U.S. Pat. No. 7,622,281 describe methods of labeling nucleic acid molecules with barcodes for the purpose of identifying the source of the nucleic acid molecules, thereby allowing for high-throughput sequencing of multiple samples (4,5). Eid et al. (2009) describe a single molecule sequencing method wherein single-molecule real time sequencing data is obtained from a DNA polymerase performing uninterrupted template-directed synthesis using four distinguishable fluorescently labeled dNTPs (6). However, these methods do not provide genomic information unaffected by amplification distortion. Miner et al. (2004) describe a method of molecular barcoding to label template DNA prior to PCR amplification, and report that the method allows for the identification of contaminant and redundant sequences by counting only distinctly tagged sequences (22). U.S. Pat. No. 7,537,897 describes methods for molecular counting by labeling molecules of an input sample with unique

oligonucleotide tags and subsequently amplifying and counting the number of different tags (23). Miner et al. and U.S. Pat. No. 7,537,897 both describe labeling of input nucleic acid molecules by ligation, which has been found to be an inefficient reaction. McCloskey et al. (2007) describe a method of molecular encoding which does not use ligation but instead uses template specific primers to barcode template DNA molecules prior to PCR amplification (24). However, such a method requires that template specific primers be made for each species of template DNA molecule studied.

Ex. 1001 at 2:1-28.

144. Below, I discuss some of these and other references that describe the use of tagging for molecular counting in conjunction with amplification. Specifically, I discuss the Brenner, Chee, McCloskey, and Porreca references, which are references that I summarize above and that reflect the general knowledge of the skilled artisan with respect to the use of tagging for molecular counting.

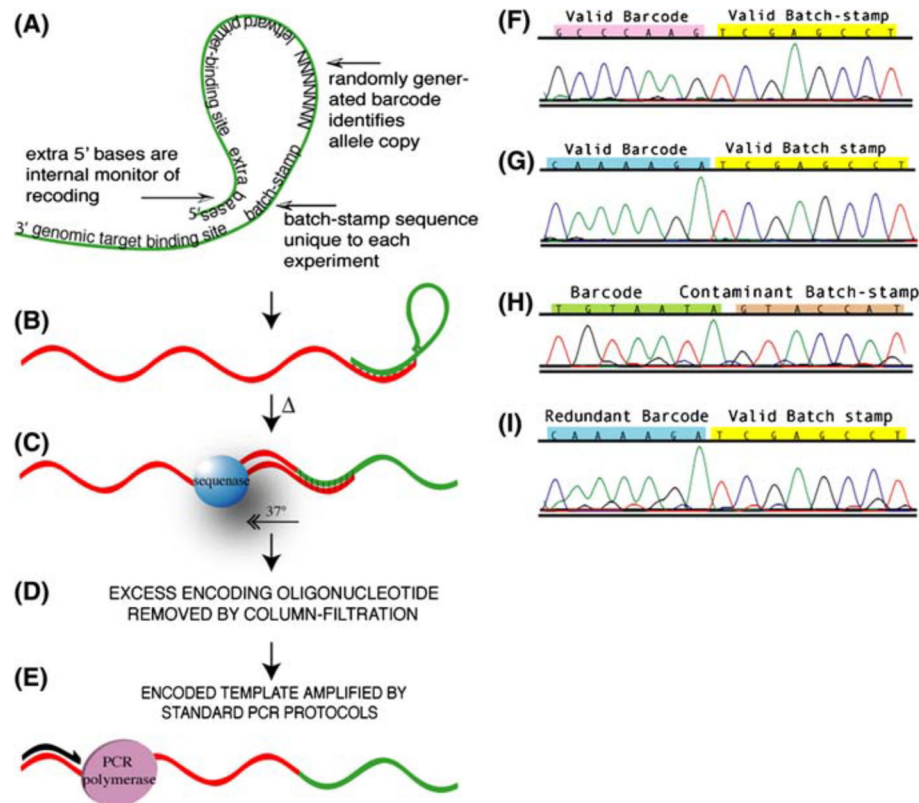
145. As I explain, each of these references disclose “tagging the segments with nucleic acid tags, thereby generating a plurality of unique tagged nucleic acid molecules, such that each of the unique tagged nucleic acid molecules comprises a segment of the segments of the genomic nucleic acids from step (a) and a nucleic acid tag of the nucleic acid tags.” I discuss this above in my overview of these references, *see supra* §§ XI.B.2, XI.C.2, XI.D.2, XI.E.2, and I rely upon that

disclosure for this claim element. Below, I discuss further relevant disclosure in these references.

146. Consider first McCloskey, which teaches using an “encoding oligonucleotide with five distinct informational regions (FIG. 1): (1) the batch-stamp, unique to each experiment, specifies the DNA source such as the patient or sample identification, and the date of reaction; (2) *the random barcode distinguishes among sequences arising from different cell or allele copies*; (3) the primer-binding site facilitates specific PCR amplification; (4) the genomic target-binding site targets the locus to be encoded prior to amplification; and (5) the extra 5’ bases serve as an internal sentinel of recoding.” Ex. 1008 at 762. The “random barcode” listed as component (2) above is the same type of tag in the ’589 patent that distinguishes among molecules and allows one to count molecules accurately after amplification. This is what McCloskey is referring to when it mentions using the barcodes to “detect template redundancy” and being able to measure “template diversity.” *Id.* at Abstract. Allowing one to distinguish sequences from different “cell or allele copies,” the introduction of the “random barcode” creates the claimed “unique tagged nucleic acid molecules.”

147. FIG. 1 of McCloskey (replicated below) shows the five distinct regions of the encoding nucleotide, including the “randomly generated barcode [that]

identifies allele copy.” *Id.* at 763. The claimed “unique tagged nucleic acid molecules” are shown beginning in step (c) below:



148. As another example, consider Porreca, which discloses a concept little different from that of the barcode in McCloskey wherein one uses “differentiator tag sequences to *uniquely tag individual nucleic acid molecules.*” Ex. 1017 at 5:3-5.

149. Porreca teaches using, among other things, “molecular inversion probe” (MIP) technology to reduce bias associated with genomic capture and sequencing, especially during detection of a “plurality of genetic loci in parallel.” *Id.* at 4:15, 3:4. Porreca’s probes are synthesized to contain “differentiator tag sequences” to “uniquely tag individual nucleic acid molecules.” i.e., to “permit the

detection of bias based on the occurrence of combinations of differentiator tag and target sequences observed in a sequencing reaction.” *Id.* at 5:3-8, 23:63-67. In this technique, the “number of each differentiator tag sequence is counted, for example, by next-generation sequencing.” *Id.* at 60:65-67; *see also id.* at 61:6-10 (describing a “sequencing technique ...used to sequence one or more captured targets (e.g., or amplicons thereof) and the sequences are used to count the number of barcodes that are present.”). Regardless, given Porreca’s teaching of using “differentiator tag sequences to ***uniquely tag individual nucleic acid molecules***,” *Id.* at 5:3-5, Porreca teaches the claimed “unique tagged nucleic acid molecules.”

150. Chee also teaches “the use of nucleic acid sequences as the unique identifiers” to tag and study “one or more desired sequences” of interest. Ex. 1015 at 11:12-14, 2:9. In general Chee teaches the concept of creating “regions of interest ***coupled with unique identifiers***.” *Id.* at 23:2-3.

151. Chee uses the unique tagged sequences generated to identify bias in the final count of sequences or Regions of Interest (ROIs). In Chee, the “methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier.” *Id.* at 10:42-46. Regardless, in teaching “regions of interest

coupled with unique identifiers,” Chee also teaches the claimed “unique tagged nucleic acid molecules.”

152. Finally, Brenner explicitly teaches an approach that provides “methods and compositions for *counting molecules in a sample, wherein each molecule is labeled with a unique oligonucleotide tag.*” Ex. 1014 at Abstract. These tags are then used for molecular counting. *See, e.g., id.* at 5:16-23 (stating that the “invention provides a method of counting molecules that are uniquely labeled with tags...substantially every molecule to be counted in a sample, e.g. the number of single stranded DNA molecules of a particular genetic locus in a sample of genomic DNA, is associated with a probe having a different tag, so that the process of counting multiple copies of the same molecule is transformed into a process of counting the number of different kinds of associated tags”); *id.* at 9:23-24 (disclosing a process wherein “binary tags are counted in each band to obtain an estimate of fragment copy number in the genomic DNA.”). In teaching molecules that are “each labeled with a unique oligonucleotide tag,” Brenner again teaches the claimed “unique tagged nucleic acid molecules.”

153. Thus McCloskey, Porreca, Chee, and Brenner each disclose “tagging the segments with nucleic acid tags, thereby generating a plurality of unique tagged nucleic acid molecules, such that each of the unique tagged nucleic acid molecules comprises a segment of the segments of the genomic nucleic acids from step (a) and

a nucleic acid tag of the nucleic acid tags.” As reflected in each of these four references, this was a standard technique that was part of the skilled artisan’s toolkit.

154. A POSA would be motivated to combine the counting approach in Lo with the generally known method of tagging (as reflected in McCloskey, Brenner, Chee and Porreca). In addition to the general discussion of why a POSA would be motivated to combine tagging and amplification with Lo, *see supra* ¶¶ 135-136, here I detail several reasons for this.

155. First, the method of tagging prior to amplification and counting as described above in §§ X.B-D and ¶¶ 144-153 was well-known and was part of the standard tool kit of a POSA, since it offered a known solution to a known problem: the need to improve the accuracy of counting molecules, particularly in situations where there were few sample DNA molecules and/or it is important to be more accurate in counting. A POSA would have been motivated to use the tagging set forth in Chee, McCloskey, Brenner, and Porreca for the simple reason that this technique represented one of only a small number of options for improving the accuracy of counting and was well within the technical grasp of a skilled artisan. Indeed, I am not aware of any other method that was utilized for improving the accuracy of counting DNA molecules in the relevant time frame.

156. Second, adding the tagging feature to Lo would be of particularly high interest to a POSA because Lo is an approach ***based on*** counting, and a POSA would

desire to count as accurately as possible. Using the tagging with Lo would enable one to subsequently use amplification without worry about amplification bias (which could be addressed using the tags). The ability to use amplification would allow one to ensure that one better captures the full range of molecules from the chromosome or chromosome region of interest, which would further improve the overall counting accuracy. Simply put, a POSA, recognizing that Lo is based on molecular counting, would naturally think to include tagging so as to perform Lo's counting method in as accurate of a manner as possible.

157. Third, and consistent with the foregoing, multiple of the references I rely upon expressly contemplate using the tagging method specifically to determine copy number information. For example, Brenner discusses using tagging and subsequent amplification to for the “measurement of genetic copy number variation.” Ex. 1014 at 4:6-7. Porreca discloses a method for “detecting nucleic acid deletions or insertions in regions containing nucleic acid sequence repeats” for cancer detection “which has been associated with micro satellite instability (MSI) involving an increase or decrease in the genomic copy number of nucleic acid repeats at one or more microsatellite loci.” Ex. 1017 at 55:48-51, 56:54-58. Yet additional prior art references expressly contemplate using tagging and amplification with counting for determining copy number information. *See* Ex. 1021 at Abstract, 1:55-2:6, 21:62-65, 5:56-6:46; *see also id.* at 18:30-49 (explaining that the disclosed

method can be used to “increase the confidence in calling copy number variations”); *id.* at 21:62-65; Ex. 1022¹ ¶ 2 (“determining the frequency of copies of gene segments for karyotyping is an important diagnostic option for diseases such as cancer); *id.* ¶ 13 (“The task is solved by labeling all nucleic acids of a nucleic acid mixture, regardless of whether they are double-stranded or single-stranded, with a ‘random tag’ (RAT) derived from a mixture of synthetic oligonucleotides, wherein the mixture contains sufficient RATs with different sequences to ensure that the resulting random tag nucleic acid (RAT-NS) combinations are practically unique.”). Thus, at least four prior art references provide express teaching to include tagging with an approach for determining copy number information based on molecular counting, as is taught in Lo.

158. Fourth, adding the tagging features described in McCloskey, Brenner, Porreca, and Chee would help solve the signal-to-noise challenge described in Lo. Specifically, Lo explains that “fetal DNA is present in low absolute concentration” in maternal plasma, creating a low signal problem. *See* Ex. 1010 ¶ 13. Additionally, fetal DNA “represent[s] a minor portion of all DNA sequences in maternal plasma and serum,” meaning it is diluted within a large background of maternal DNA, which contributes to high noise in the sample. *See id.* A POSA would be motivated to

¹ Certified translation of German original Ex. 1040. *See* Ex. 1041.

combine the tagging techniques known in the art with the amplification techniques described in step [1C] to increase the signal-to-noise ratio and thereby improve the sensitivity and accuracy of detecting aneuploidy described in Lo. Lo emphasized that it was “desirable to have methods that allow the noninvasive detection of fetal chromosomal aneuploidy by maximizing the amount of genetic information that could be inferred from the limited amount of fetal nucleic acids which exist as a minor population in a biological sample containing maternal background nucleic acids.” *Id.*

159. A POSA would have had a reasonable expectation of successfully making this combination because the machinery for preparing and attaching the tags was well-known and straightforward.

160. A POSA would know, for instance, that it is straightforward to design a large number of random sequences to achieve the claimed uniqueness. Brenner, for instance, describes using oligonucleotide tagging sets which are “preferably at least ten times the number of target polynucleotides to be counted,” and “[w]here labeling by sampling is employed, preferably, the size of sets of oligonucleotide tags of the invention are at least 100, and in another aspect, at least 1000; and in still another aspect at least 10,000.” Ex. 1014 at 25:36-46. Brenner similarly contemplates that the tags should be “maximally different,” *id.* at 21:14-18, and teaches that there were well known ways of making the tags “maximally different.”

Brenner explains, for instance, that there were “distance measures” known in the art for designing such sequences. Brenner cites several literature references to this effect, stating “exemplary distance measures for sequences applicable to the invention are well known and are widely disclosed in the literature, as exemplified by the following references: Gusfield, Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology (Cambridge University Press, 1997); Navarro et al, Flexible Pattern Matching on Strings (Cambridge University Press, 2002); Sankoff et al, editors, Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison (Center for the Study of Languages, 1999); and the like.” *Id.* at 25:58-66.

161. Chee likewise uses a “large number of unique sequences that can be generated from the random assembly of nucleotides [which] provides a high probability that each individual construct will be uniquely identified, even those constructs that contain substantially the same specific ROIs. The number of nucleotides in the identifier is preferably chosen so that the number of potential and actual sequences represented is much greater than the total number of oligonucleotide constructs in the set.” Ex. 1015 at 11:26-38. And Porreca similarly discloses a method of creating a “large library of unique differentiator tag sequences...using degenerate, random-sequence polynucleotides of defined length.”

Ex. 1017 at 29:37-39. Thus, methods were available for preparing a large number of different oligonucleotide tags.

162. So too were methods for attaching the tags. For example, Porreca describes various methods for combining tags with target sequences. Consider, for instance, FIG. 4A-C of Porreca below:

FIG. 4A

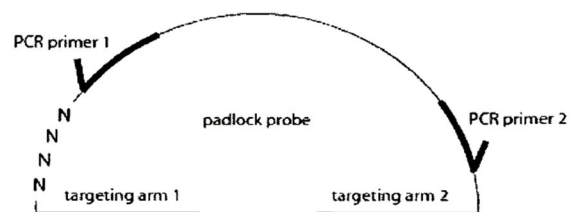


FIG. 4B

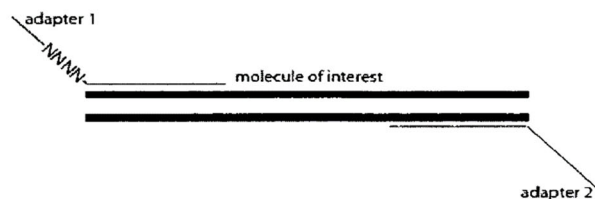
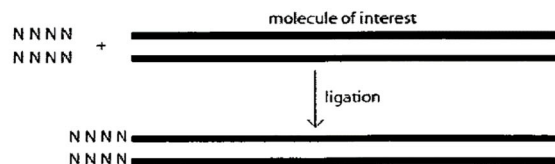


FIG. 4C



Id. at FIG. 4A-4C.

163. Brenner likewise details a general procedure for attaching an oligonucleotide tag to one end of a polynucleotide (see FIGS. 2A-2B below), including methods for attachment chemically or enzymatically. Consider for

instance Brenner, Examples I and II, which describe the construction of metric and binary tag sets, respectively:

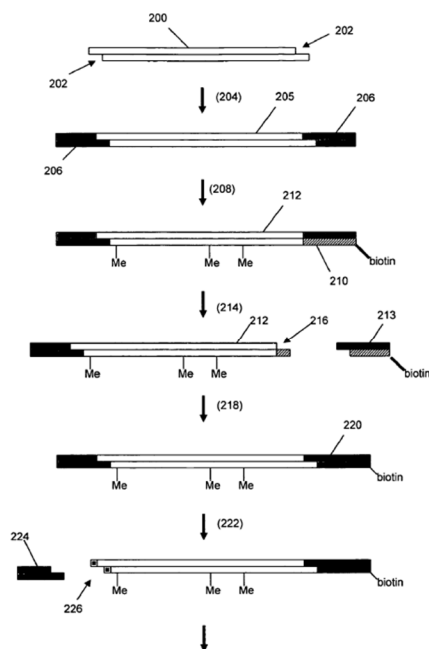


Fig. 2A

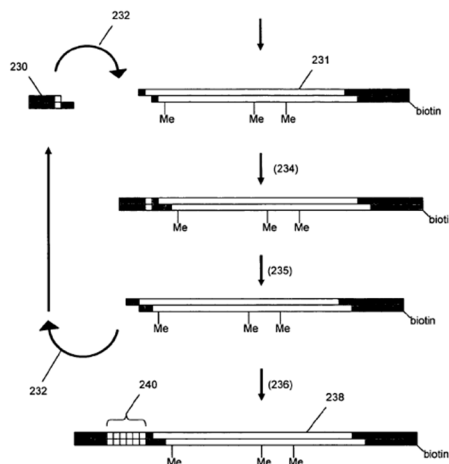


Fig. 2B

Ex. 1014 at FIG. 2A-2B.

164. In any event, there can be no question of reasonable expectation of success here because, as I explain above, the '589 patent admits that effective methods of tagging existed in the art. *See* Ex. 1001 at 2:20-22 (describing labeling by ligation); *id.* at 2:23-26 (describing labeling through specific primers); *id.* at 21:38-40 (“Tags may be associated with nucleic acid molecules by using, singly or in combination, ligation, nick-ligation, terminal transferase, hybridization, priming, and the like.”).

165. In sum, it is my opinion that a POSA would be motivated to combine the well-known tagging of McCloskey, Brenner, Chee, or Porreca, with the counting method of Lo to increase accuracy in the final genomic fragment counts. Likewise, it is my opinion that a POSA would have a reasonable expectation of success, based on well-known methods of tagging as practiced in the cited references herein.

c) [1C] subjecting the plurality of the unique tagged nucleic acid molecules to a polymerase chain reaction (PCR), thereby generating amplified tagged nucleic acid molecules;

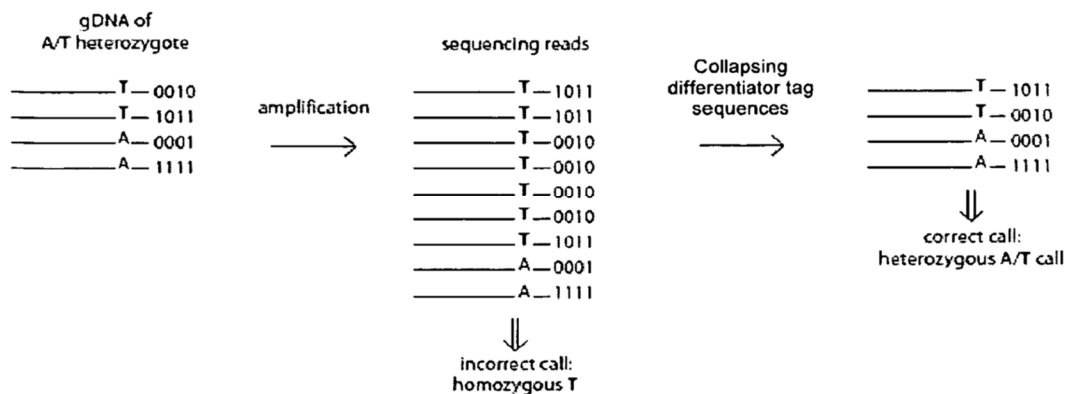
166. As discussed immediately above in connection with element [1B], it is my opinion that it would have been obvious to tag the DNA fragments utilized in Lo's method to improve the accuracy of molecular counting. This technique (as disclosed in Brenner, Porreca, Chee, and McCloskey) is a method that would have been within the common knowledge of the skilled artisan. The skilled artisan would have further understood that this approach includes amplification of the tagged fragments. Indeed, this is disclosed in each of Brenner, Porreca, Chee, and McCloskey. I discuss this above in my overview of these references, *see supra* §§ XI.B.2, XI.C.2, XI.D.2, XI.E.2, and I rely upon that disclosure for this claim element. Below, I discuss further relevant disclosure in these references.

167. As discussed above, McCloskey teaches tagging with an encoding nucleotide containing five distinct regions. In a specific example involving the FMR1 locus, for instance, McCloskey tags the relevant DNA molecules after which

“standard PCR amplification to the human locus FMR1” was applied. Ex. 1008 at 766. Thus, McCloskey discloses PCR amplification of unique tagged nucleic acid molecules.

168. Porreca likewise teaches the use of PCR methods that “may include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel). Ex. 1017 at 16:12-15. FIG. 5 of Porreca (see below) shows a method for amplification of tagged target DNA fragments, followed by counting of the tagged reads to reduce amplification bias.

FIG. 5



Id. at FIG. 5; *see also id.* at 15:64-67-16:1-2 (describing that in “amplification-based (e.g., PCR-based or LCR-based, etc.) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction or ligase

chain (or other amplification method) that selectively amplifies each locus using a pair of oligonucleotide primers.”); *id.* at 58:43-49 (“PCR...is performed to isolate and amplify the circular targets from the starting nucleic acid pool.”); *id.* at 58:50-53 (“repetitive regions are surrounded by non-repetitive unique sequences, which can be used to amplify the repeat-containing regions using for example, PCR.”). Thus, Porreca further discloses PCR amplification of unique tagged nucleic acid molecules.

169. Brenner describes a technique for amplification after sequence tagging wherein, for example, “*oligonucleotide tags may be amplified using emulsion PCR, e.g. as also disclosed in the cited references*, so that clonal populations of each oligonucleotide tag of a sample are formed on beads, *which are then sequenced.*” Ex. 1014 at 17:53-55. Brenner emphasizes in the disclosure of the method of metric tagging to adjust the tagging lengths to streamline the subsequent amplification, stating that although “each genome is in a one-to-one correspondence with a single length of an oligonucleotide (i.e. a metric tag), the physical lengths of the metric tags are not the same and *since it is desirable to be able to PCR the tags* preferably the metric tags should be the same length. Thus, appropriate length of oligonucleotide are added to each to make them all the same.” *Id.* at 17:14-20. Additionally, Example 1 of Brenner describes a detailed technique using PCR to amplify 64 plasmid DNA samples containing tags. *Id.* at 29:27-54.

170. Finally, Chee describes “using techniques such as PCR or emulsion PCR” for amplification in a solution of target sequences. Ex. 1015 at 13:40-41. Chee teaches the reason or motivation for choosing this technique, that “methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier.” *Id.* at 10:41-46. Chee further discloses that to “truly minimize the amount of amplification needed, the labeling and sorting of individual molecules can be employed prior to the amplification step in the present methods.” *Id.* at 15:64-66. Thus, Chee further discloses PCR amplification of unique tagged nucleic acid molecules.

171. A POSA would have been motivated to use amplification with tagging for several reasons. Most important, amplification is part and parcel of the methods set forth in Brenner, McCloskey, Chee, and Porreca, which is all reflective of the knowledge of the skilled artisan. Indeed, there is no reason to utilize the tagging methods described in these prior art references without subsequent amplification because the whole point of the tagging is to permit one to subsequently perform amplification and hence ensure that one captures all the molecules of interest while being able to correct for amplification bias. Thus, all the reasons set forth above in

connection with element [1B] as to why a POSA would be motivated to apply tagging to Lo's method apply equally to the use of amplification.

172. Further, exemplary samples in Lo's method comprised cell-free DNA in bodily fluids from a subject, such as blood or urine samples. Ex. 1010 ¶ 14. As a POSA would understand, and as Lo confirms, with such samples there is a limited amount of DNA overall and even less DNA from the fetus that one is interested in testing. *See id.* ¶¶ 13, 16, 18. Therefore, a POSA would have naturally been motivated to perform amplification with Lo's methods to increase the overall amount of DNA for testing.

173. Yet additional motivation to utilize amplification is present within Lo itself. Specifically, Lo teaches multiple methods of enriching nucleic acid sequences prior to sequencing and the motivation for doing so, namely for cost-saving and efficiency:

- Lo teaches that in “some embodiments of the present invention, the random sequencing may be preceded by procedures to enrich a biological sample with particular populations of nucleic acid molecules sharing certain common features.” *Id.* ¶ 47.
- Lo explains that it would be “possible and cost-effective to enrich the pool of nucleic acids to be sequenced prior to random sequencing of a fraction of the enriched pool.” *Id.* ¶ 77.

- Lo teaches that since the fetal DNA found in the maternal plasma is generally of shorter length than the maternal chromosomal fragments, one method to accomplish enrichment would be to “fractionate the nucleic acid sequences in the sample according to molecule size, e.g. by gel electrophoresis or size exclusion columns or by microfluidics-based approach.” *Id.* .
- Lo also teaches the use of hybridization techniques to accomplish enrichment of potentially aneuploid chromosomes and one or more normal chromosomes “for example onto oligonucleotide microarrays,” followed by random sequencing. *Id.* ¶ 79. This is a known cost-saving technique to those skilled in the art.
- Lo describes that pre-sequencing enrichment in fetal cell free DNA could be accomplished through suppression of the maternal DNA relative to fetal DNA with the use of chemical treatments such as formaldehyde. *Id.* ¶ 72.
- Lo teaches the concept of sub-selecting some desired sequences from the pool of nucleic acids prior to sequencing. “For example, hybridization based techniques such as oligonucleotide array could be used to first sub-select for nucleic acid sequences from certain chromosomes, e.g. a potentially aneuploid chromosome and other

chromosome(s) not involved in the aneuploidy tested. Another example is that a certain sub-population of nucleic acid sequences from the sample pool is sub-selected or enriched prior to sequencing.” *Id.*

174. The enrichment taught by Lo is an improvement over previous whole-genome sequencing methods and is well-known to those of skill in the art, providing enhanced efficiency and cost-savings, as described above. The amplification as taught by McCloskey, Brenner, Porreca or Chee is another form of enrichment, and therefore provides the same benefit but with the additional increase in counting accuracy by virtue of the tagging.

175. In my opinion, a POSA would be motivated to combine the amplification and tagging of McCloskey, Brenner, Chee, and Porreca, with the enrichment method of Lo to increase efficiency and cost savings without the loss of accuracy in the final genomic fragment counts. Indeed, Lo expressly contemplates enrichment, and PCR is a well-established and routine method to provide enrichment.

176. Further, a POSA would have a reasonable expectation of success because there is no question that PCR amplification was routine and widely used well before the priority date. Indeed, each of Brenner, Chee, Porreca and McCloskey disclose the use of PCR. And the '589 patent effectively admits this that PCR was a well-known and routine technique. Specifically, the specification explains that

“[g]enerally...[the amplification] steps are performed multiple times (e.g., polymerase chain reaction).” Ex. 1010 at 13:50-54; *see also id.* at 1:37-2:28 (describing prior art methods employing PCR). The Abstract also references “subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction” *Id.* at Abstract. In other words, the ’589 patent rightly characterizes PCR as a well-known prior art method that people of skill in the art had no trouble utilizing.

d) [1D] generating tag associated sequence reads by sequencing the amplified tagged nucleic acid molecules of step (c);

177. Lo in combination with the knowledge of the skilled artisan (as reflected, for example, in Brenner, McCloskey, Chee, and Porreca) renders obvious “generating tag associated sequence reads by sequencing the amplified tagged nucleic acid molecules of step (c).” As discussed above in connection with claim elements [1B] and [1C], a POSA would have been motivated to utilize the tagging and amplification of sequences in connection with Lo’s method for the detection of copy number information. Having been motivated to use tagging and amplification with a reasonable expectation of success, a skilled artisan would have further proceeded according to Lo’s standard method, which involves sequencing of nucleic acid molecules for the purpose of determining copy number information.

178. Lo describes a process, wherein in “step 120, at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced. The portion sequenced represents a fraction of the human genome. In one embodiment, the nucleic acid molecules are fragments of respective chromosomes. One end (e.g. 35 basepairs (bp)), both ends, or the entire fragment may be sequenced. All of the nucleic acid molecules in the sample may be sequenced, or just a subset may be sequenced. This subset may be randomly chosen.” *Id.* ¶ 55.

179. FIG. 6 of Lo (reproduced below) shows the result of sequencing the DNA fragments (sequence tags) generated from each of the tested maternal plasma specimens as described in Experiment I.

Sample	Total no. of sequenced tags	Total no. of sequenced basepairs	Proportion of the human genome (%)	U0 count	U0 count / total count (%)	No. of U0 sequenced basepairs	U0 Proportion of the human genome (%)
Sample 1 T21	1.12E+07	4.02E+08	13.4	1.93E+06	17.3	6.94E+07	2.31
Sample 2 T21	1.04E+07	3.73E+08	12.4	1.80E+06	17.4	6.50E+07	2.17
Sample 3 T21	8.90E+06	3.20E+08	10.7	2.09E+06	23.4	7.51E+07	2.50
Sample 4 T21	1.02E+07	3.69E+08	12.3	2.23E+06	21.7	8.02E+07	2.67
Sample 5 euploid	1.06E+07	3.83E+08	12.8	2.12E+06	19.9	7.62E+07	2.54
Sample 6 euploid	9.58E+06	3.45E+08	11.5	1.91E+06	20.0	6.88E+07	2.29
Sample 7 euploid	9.55E+06	3.44E+08	11.5	2.01E+06	21.0	7.22E+07	2.41
Sample 8 euploid	9.09E+06	3.27E+08	10.9	2.09E+06	23.0	7.53E+07	2.51

FIG. 6

Id. at FIG. 6.

180. Experiment II explains that “only the uniquely mappable sequenced tags, termed U0 in nomenclature from the Eland software, were used to demonstrate the presence of over-representation in the amount of chromosome 21 sequences in the maternal plasma specimens from pregnancies each carrying a fetus with trisomy 21, as described in example I. As shown in FIG. 6, ***U0 sequences only represent a subset of all the sequenced tags generated from each specimen and further represent an even smaller proportion***, some 2%, of the human genome.” *Id.* ¶ 112; *see also id.* ¶¶ 56, 73 (disclosing single molecule sequencing technologies including “the Roche 454 platform, the Applied Biosystems SOLiD platform, the the [sic] Helicos True Single Molecule DNA sequencing technology, the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and nanopore sequencing”).

181. The only difference between this claim step and the method set forth in Lo is that this claim step involves sequencing tagged and amplified nucleic acid molecules whereas Lo directly sequences the original sample molecules. Yet, as set forth above in connection with elements [1B] and [1C], a skilled artisan would have found it obvious to include tagging and amplification with Lo’s method. Having been motivated to modify Lo’s method to use tagging and amplification, a POSA would have necessarily sequenced the tagged and amplified molecules, according to the methods and techniques set forth in Lo that I summarize above.

182. In sum, it is my opinion that a POSA would have been motivated to combine the tagging and amplification of McCloskey, Brenner, Chee, and Porreca, as described above in claim elements [1B] and [1C] with the sequencing method of Lo, to increase the accuracy of the counting method, for the purpose of determining copy number information. Likewise, it is my opinion that a POSA would have had a reasonable expectation of success, given how well-known the methods of sequencing were as practiced in the cited references herein. Indeed, as noted above, Lo proposes to use well-known commercial available sequencing platforms, such as Roch 454, SOLiD, Helios, and Pacific Biosciences, similar to the disclosure of the '589 specification.

- e) **[1E] assigning each of the amplified tagged nucleic acid molecules to a location on a reference genome from the same species from which the sample has been obtained by mapping each of the tag associated sequence read of step (d) to a location of the reference genome; and**

183. In my opinion, a POSA would have been motivated to perform this claim step in conjunction with Lo's method for determining copy number information.

184. To begin, Lo teaches the process of assigning nucleic acid sequences to a reference by mapping. Lo teaches, for instance, that "based on the sequencing (e.g. data from the sequencing), a first amount of a first chromosome (e.g. the clinically relevant chromosome) is determined. The first amount is determined from

sequences identified as originating from the first chromosome. For example, a bioinformatics procedure may then be used to locate each of these DNA sequences to the human genome.” *Id.* ¶ 58.

185. Further, Lo teaches that “[e]ach genomic DNA or plasma DNA sample was sequenced using one lane of a flow cell. The short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted. The total number of individual sequenced tags aligned to each chromosome were tabulated and compared with the relative size of each chromosome as expected from the reference human genome or non-disease representative specimens.” *Id.* ¶ 70.

186. The Lo reference provides, in Example I, details regarding a study wherein maternal plasma DNA was collected from women with aneuploid and euploid pregnancies. After DNA from each maternal plasma specimen was sequenced, “[s]equenced reads were compiled using Solexa Analysis Pipeline. All reads were then aligned to the repeat-masked reference human genomic sequence, NCBI 36 assembly.” *Id.* ¶ 93.

187. Figure 4A of Lo (reproduced below) shows the percentages of maternal plasma DNA aligned to each of the 24 chromosomes (22 autosomes and X and Y chromosomes) for two representative cases, one with a euploid fetus and the other carrying a trisomy-21 fetus:

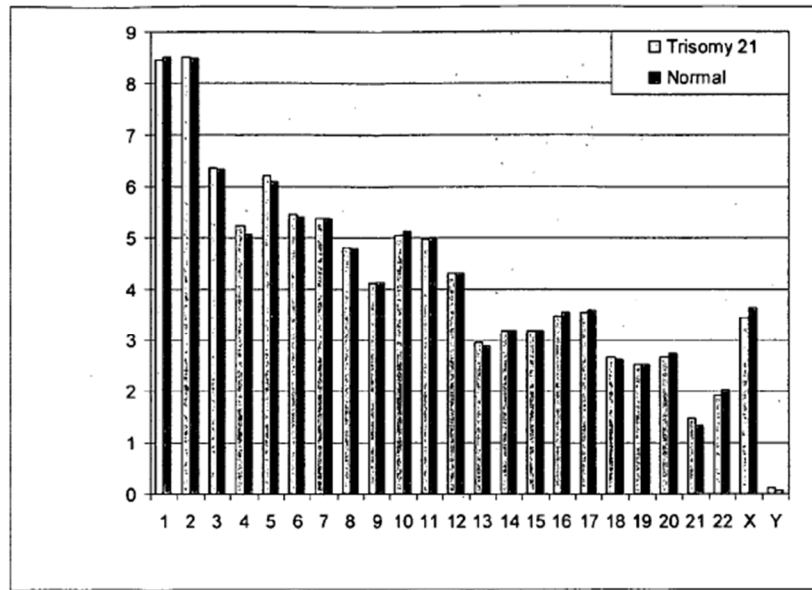


FIG. 4A

Id. at FIG. 4A.

188. The only difference between this '589 claim step and the method set forth in Lo is that this claim step involves mapping tagged and amplified nucleic acid molecules whereas Lo aligns the sequences of the original sample molecules. Yet, as set forth above in connection with elements [1B] and [1C], a skilled artisan would have found it obvious to include tagging and amplification with Lo's method. Having been motivated to modify Lo's method to use tagging and amplification, a POSA would have necessarily sequenced and then aligned the sequences of the tagged and amplified molecules, according to the methods and techniques set forth in Lo that I summarize above. There can be no question that a POSA would have had a reasonable expectation of successfully doing this given that alignment methods were well-known and routine. Lo even reports actual data showing that it was easily

achievable to align sequence reads and use the aligned sequence reads to determine copy number information.

- f) **[1F] at a plurality of locations on the reference genome, counting the number of unique tagged nucleic acid molecules that have been assigned to the same location on the reference genome based on the mapping step to obtain a count for each location of the plurality of locations on the reference genome, thereby obtaining copy number information from genomic nucleic acids in the sample.**

189. In my opinion, in view of the knowledge of the skilled artisan, it would have been obvious to perform this claim step in conjunction with Lo's method. As I explain above, in connection with steps [1B] and [1C], a POSA would have found it obvious to use tagging and amplification in conjunction with Lo's method, which, as I explain immediately below, is based on counting molecules.

190. To begin, Lo is clear that its method is based on counting molecules to obtain copy number information. Lo discloses, for instance, that there "are a number of ways of determining the amounts of the chromosomes, including but not limited to *counting* the number of sequenced tags, the number of sequenced nucleotides (basepairs) or the accumulated lengths of sequenced nucleotides (basepairs) originating from particular chromosome(s) or chromosomal regions." *Id.* ¶ 60.

191. Lo then discloses how to utilize the sequence tags that were counted as described in paragraph 60 to derive copy number information. Specifically, Lo teaches that the "number of sequenced tags, i.e. absolute count, aligned to a

particular chromosome could be compared to the absolute count of sequenced tags aligned to other chromosomes. Alternatively, the fractional count of the amount of sequenced tags from chromosome 21 with reference to all or some other sequenced tags could be compared to that of other non-aneuploid chromosomes. In the present experiment, because 36 bp were sequenced from each DNA fragment, the number of nucleotides sequenced from a particular chromosome could easily be derived from 36 bp multiplied by the sequenced tag count.” *Id.* ¶ 75.

192. Lo further discloses how to utilize the counted sequence tags to derive yet additional copy number information, specifically an estimation of copy number. Lo teaches, for instance considering a “random representative fraction of DNA molecules from the specimen pool” and then comparing the “number of different sequenced tags aligned to various chromosomal regions...between specimens containing or not containing the DNA species of interest. Chromosomal aberrations would be revealed by differences in the number (or percentage) of sequences aligned to any given chromosomal region in the specimens.” *Id.* ¶ 108. Lo explains that any type of clinically relevant sequence that one may identify by these means are “copy number variations.” *Id.* ¶ 37. Lo teaches the mathematical formula for doing this analysis and includes actual experimental data showing that his method works. *See id.* ¶¶ 95-96, 99-104; *id.* at FIGs. 3A, 4B, 5.

193. Thus, there is no question that Lo’s method is based on counting nucleic acid molecules to estimate copy number. Moreover, Lo teaches that the counting for this purpose can be done at a “plurality of locations” in the reference genome. Specifically, Lo teaches that the analysis can be done at a plurality of chromosomes or a plurality of regions in the genome. *See, e.g., id.* ¶ 109 (“Changes (amplifications and deletions) in multiple chromosomal regions may be used. Thus, there would be an increased proportion of sequences aligned to the amplified regions and a decreased proportion of sequences aligned to decreased regions.”); *id.* at claim 5 (“5. The method of claim 1, wherein the first chromosome is chromosome 21, chromosome 18, chromosome 13, chromosome X, or chromosome Y”); *id.* ¶ 14 (“One or more cutoff values are chosen for determining whether a change compared to a reference quantity exists (i.e. an imbalance), for example, with regards to the ratio of amounts of two chromosomal regions (or sets of regions).”); *id.* ¶ 67 (“Also, dosage imbalance of a particular chromosome or chromosomal regions can be quantitatively determined. In other words, the dosage imbalance of the chromosome or chromosomal regions is inferred from the percentage representation of the said locus among other mappable sequenced tags of the specimen.”); *id.* ¶ 71 (“Gains or losses of chromosomal regions or whole chromosomes were determined by comparing the tag counts with the expected chromosome size in the reference genome or that of a non-disease representative specimen.”); *id.* ¶ 108 (“The number

of different sequenced tags aligned to various chromosomal regions is compared between specimens containing or not containing the DNA species of interest. Chromosomal aberrations would be revealed by differences in the number (or percentage) of sequences aligned to any given chromosomal region in the specimens.”).

194. Lo’s mapping to a “chromosomal region” is consistent with PO’s reading of “location” in the parallel district court litigation. See Ex. 1020.049-.050 (interpreting locations as any of “genes,” “genetic locus,” and “start/stop positions”). This count of molecules in a region comprises “copy number information.”

195. The only difference between this claim step and the method set forth in Lo is that this claim step involves counting “unique tagged nucleic acid molecules,” whereas Lo counts original sample molecules. Yet, as set forth above in connection with elements [1B] and [1C], a skilled artisan would have found it obvious to include tagging and amplification with Lo’s method. Having been motivated to modify Lo’s method to use tagging and amplification, a POSA would have naturally counted the “unique” molecules. Indeed, as I explain above in connection with steps [1B] and [1C], the main reason for using the tagging and amplification in the first place is to increase the accuracy of the counting, which is accomplished by employing the tags to identify and count only the unique molecules, as is done in each of Chee, Brenner, McCloskey, and Porreca. To count anything other than the unique molecules would

defeat the purpose of using the tagging and amplification that the skilled artisan would have been motivated to use.

196. In this regard, even in Lo's method, he proposed to only count the molecules that could be uniquely mapped to a single, specific location. *See, e.g.*, Ex. 1010 ¶ 85 ("For example, in example II, sequenced tags that were uniquely aligned to the repeat-masked reference human genome were used."); *id.* ¶ 112 ("Furthermore, in this study, only the uniquely mappable sequenced tags, termed U0 in nomenclature from the Eland software, were used...."). In this way, Lo avoided including sequence reads that might distort the count of sequence reads mapped to a particular region. A skilled artisan using tagging and amplification in Lo's method would have acted consistently with Lo's interest in having accurate sequence counts and used only the unique sequence reads rather than counting duplicates and hence overinflating read counts.

197. Therefore, in my opinion, a POSA would have found it obvious to perform this claim step in view of the Lo disclosure.

2. Claim 2: The method of claim 1, further comprising estimating a copy number at multiple locations within a region of the reference genome.

198. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 2 or, at a minimum, claim 2 would have been obvious in view of Lo.

199. As I explain above in connection with claim 1, Lo teaches that its counting can be done at multiple regions of a chromosome. *See, e.g., Id.* ¶ 14 (“One or more cutoff values are chosen for determining whether a change compared to a reference quantity exists (i.e. an imbalance), for example, with regards to the ratio of amounts of two chromosomal regions (or sets of regions).”); *id.* ¶ 67 (“Also, dosage imbalance of a particular chromosome or chromosomal regions can be quantitatively determined. In other words, the dosage imbalance of the chromosome or chromosomal regions is inferred from the percentage representation of the said locus among other mappable sequenced tags of the specimen.”); *id.* ¶ 71 (“Gains or losses of chromosomal regions or whole chromosomes were determined by comparing the tag counts with the expected chromosome size in the reference genome or that of a non-disease representative specimen.”); *id.* ¶ 108 (“The number of different sequenced tags aligned to various chromosomal regions is compared between specimens containing or not containing the DNA species of interest. Chromosomal aberrations would be revealed by differences in the number (or percentage) of sequences aligned to any given chromosomal region in the specimens.”).

200. Using a chromosome as a region is consistent with PO’s interpretation of “region” in the parallel district court litigation as a section of the genome comprising multiple genes. *See* Ex. 1020.050 (discussing “the copy number of

different genes, which are located at different locations in a given region of the reference genome.”). To the extent Lo is determined to not disclose this claim element explicitly, a POSA would have nonetheless found it obvious in view of Lo. Indeed, as noted above, Lo explicitly contemplates analyzing *plural* “chromosomal regions.” This makes sense; Lo’s method is based on determining an imbalance in the amount of DNA arising from a region of the genome, and this approach is not limited to a particular chromosome, as is clear from the disclosure in the preceding paragraph. A POSA would have found it obvious to analyze multiple discrete regions of a particular chromosome because, for instance, there were several known copy number alterations that existed on a single chromosome that were of clinical relevance.

201. For example, Itsara *et. al* published a large-scale study as early as 2009, examining copy number variations in human genetic samples. The authors found multiple CNVs on the same chromosome, specifically in regions of the genome previously known to be associated with certain diseases. Ex. 1035. The authors correlated the frequencies and sizes of these copy number variations with known pathogenic hotspots, for further understanding of disease mechanisms. For example, multiple chromosome deletions on chr15 was correlated to a range of neurological or mental illnesses, such as schizophrenia etc. *Id. at* 11. Table 2 reproduced below shows CNV data across chromosomal locations and regions which correlated with

autism, schizophrenia and mental retardation in the Itsara study, and which include examples of multiple locations along a chromosomal region.

Table 2. Loci Enriched for CNVs in Autism, Mental Retardation, and Schizophrenia Identified by Disease Meta-Analysis

Chr	Start	Stop	Length	Note	NAHR ^a	Total CNVs	Type	Disease CNVs	Control CNVs	Diseases and Studies	CNV p Value	Locus p Value
chr15	27,015,263	30,650,000	3,634,737	Prader-Willi/15q13	yes	19	loss	19	0	schizophrenia ²³	1.08E-05	
chr15	18,376,200	30,756,771	12,380,571			58	gain	45	13	autism, ^{19,22,45-47} mental retardation, ⁴⁴ schizophrenia, ²³ controls, ²³ this study	2.69E-04	1.54E-07
chr22	17,014,900	19,993,127	2,978,227	VCFS	yes	31	loss	31	0	autism, ^{19,45,47} mental retardation, ⁴⁴ schizophrenia, ^{23,48} controls, ²³ this study	7.93E-09	
chr22	17,200,000	21,546,762	4,346,762			14	gain	9	5	autism, ^{19,45-47} schizophrenia, ²³ controls, ²³ this study	0.330	9.53E-07
chr1	142,540,000	146,059,433	3,519,433	1q21	yes	27	loss	24	3	autism, ¹⁹ schizophrenia, ²³ controls, ²³ this study	1.67E-04	
chr1	142,800,580	146,009,436	3,208,856			15	gain	12	3	autism, ^{19,45} mental retardation, ⁴⁴ schizophrenia, ²³ controls, ²³ this study	0.041	2.16E-05
chr22	45,144,027	49,509,153	4,365,126	Terminal 22 del syndrome	no	4	loss	4	0	autism ^{22,45,47}	0.090	
chr22	47,572,875	48,323,417	750,542			6	gain	5	1	autism, ^{19,46,47} schizophrenia, ²³ controls, ²³ this study	0.160	0.022
chr16	29,470,951	30,252,473	781,522	16p11.2	yes	11	loss	8	3	autism, ^{19,22,46,47} schizophrenia, ²³ controls, ²³ this study	0.186	
chr16	29,474,810	30,235,818	761,008			7	gain	6	1	autism, ^{19,47} schizophrenia, ^{23,24} controls, ²³ this study	0.100	0.039
chr17	14,000,000	15,421,835	1,421,835	CMT1A/HNPP	yes	7	loss	6	1	autism, ^{19,46} schizophrenia, ²³ controls, ²³ this study	0.100	
chr17	12,650,000	15,540,000	2,890,000			5	gain	4	1	autism, ⁴⁵ mental retardation, ⁴⁴ schizophrenia, ²³ controls, ²³ this study	0.252	0.041
chr16	60,141,700	61,581,600	1,439,900	16q21, CDH8	no	4	loss	4	0	autism ⁴⁵	0.090	
chr16	60,552,237	61,294,685	742,448			1	gain	1	0	schizophrenia ²³	0.547	0.049
chr11	78,120,000	85,610,000	7,490,000	11q14.1	no	3	loss	3	0	mental retardation, ⁴⁴ schizophrenia ²³	0.164	
chr11	84,304,683	85,042,205	737,522			1	gain	1	0	schizophrenia ²³	0.547	0.090
chr2	185,118,087	185,909,729	791,642	2q32.1	no	1	loss	1	0	schizophrenia ²³	0.547	
chr2	184,270,000	186,892,000	2,622,000			3	gain	3	0	autism ⁴⁵	0.164	0.090
chr15	82,573,421	83,631,697	1,058,276	15q25	yes	4	loss	4	0	autism, ^{46,47} schizophrenia ²³	0.090	
chr15	na	na	na			0	gain	0	0	none	1	0.090
chr9	140575	1175526	1,034,951	9p24	no	1	loss	1	0	schizophrenia, ²³ controls, ²³ this study	0.547	
chr9	206456	1599250	1,392,794			3	gain	3	0	autism, ⁴⁵ schizophrenia, ²³ controls, ²³ this study	0.164	0.090
chr3	197,179,156	198,842,299	1,663,143	3q29	yes	3	loss	3	0	schizophrenia ^{23,24}	0.164	
chr3	198,325,925	199,384,429	1,058,504			2	gain	1	1	schizophrenia, ²³ controls, ²³ this study	1	0.252
chr16	21,693,739	22,611,363	917,624	16p12	yes	5 ^b	loss	5 ^b	0	autism, ⁴⁷ schizophrenia ²³	0.049	
chr16	21,441,805	22,688,093	1,246,288			5	gain	2	3	autism, ⁴⁷ schizophrenia, ²³ controls, ²³ this study	1	0.261
chr16	80,722,684	82,227,917	1,505,233	16q23.3, CDH13	no	2	loss	0	2	controls, ²³ this study	1	
chr16	80,737,839	82,208,451	1,470,612			4	gain	4	0	autism, ⁴⁶ schizophrenia ^{23,24}	0.090	0.436

^a Indicates if there are large segmental duplications near breakpoints in hg17.

^b A deletion of ~480 kb in size in a schizophrenic sample is included in this count.

202. In yet another example, Mefford *et. al.* describe a new method for targeting a set of known pathogenic CNVs related to neurocognitive disease. The

authors examined 69 hotspot regions of the genome known to be likely candidates for deletion or duplication, including multiple examples of pathogenic CNVs on the same chromosome. Table 1 replicated below shows the CNVs identified in multiple locations in regions of the genomes and correlated to subjects with intellectual disabilities. Ex. 1036.

Table 1. CNVs confirmed or predicted in series of 1010 individuals with ID

Locus	Size (MB)	Confirmed by array-CGH	Predicted ^a	Combined frequency (n = 1010)
Pathogenic events^b				
1q21.1 dup	1.4	1	2	0.30%
15q11 BP1-BP3 del ^c	5.7	1	0	0.10%
15q11 BP2-BP3 del	5.3	1	0	0.10%
15q24 del	2.2	0	1	0.10%
16p11.2 del	0.9	6	1	0.69%
16p11.2 dup	0.9	1	0	0.10%
16p12.2-p11.2 del	8.1	0	1	0.10%
16p13 BP1-BP2 del	1.2	1	1	0.20%
16p13 BP1-BP3 del ^c	3.2	1	0	0.10%
17p11.2 del (SMS) ^c	3.5	2	0	0.20%
17p12 del (HNPP)	1.4	1	0	0.10%
17p12 dup (CMT1A)	1.4	0	1	0.10%
17q12 del	1.6	1	0	0.10%
22q11.21 3-Mb del (VCFS) ^c	2.9	2	1	0.30%
22q11.21 3-Mb dup ^c	2.9	2	0	0.20%
22q11.21-q11.22 distal del	1.4	2	1	0.30%
Total		22	9	3.08%
Possibly pathogenic events^b				
15q11.2 BP1-BP2 del	0.4	7	1	0.79%
16p13 BP1-BP2 dup	1.2	7	1	0.79%
16p13 BP1-BP3 dup ^c	3.2	2	1	0.30%
Large 16p13 dup (8Mb) ^d	15	1	0	0.10%
17q12 dup	1.6	1	0	0.10%
22q11 distal BCR dup	0.7	0	1	0.10%
VCFS distal del	0.9	1	0	0.10%
Total		19	4	2.28%
Total path + possible		41	13	5.35%
Events of unknown significance				
7q11 del (distal to WBS)	1.7	1	0	0.10%
10q11 del	2.6	1	0	0.10%
15q11.2 BP1-BP2 dup	0.4	1	1	0.20%
11q13 dup	3.9	0	1	0.10%
Polymorphic CNVs				
NPHP1 del	0.7	2	0	0.20%
NPHP1 dup	0.7	1	6	0.69%
1q21.1 (TAR region) dup	0.3	3	NA	0.30%
22q11.22-q11.23 del ^e	0.7	1	0	0.10%
Total CNVs detected		51	21	7.14%

203. As such, a POSA applying Lo's method to detect such cases would be motivated to analyze "multiple locations" within a region of a reference genome to look for copy number variations in a sample. A POSA would have had a reasonable

expectation of successfully doing this because it represents nothing more than Lo's method applied to multiple regions of a chromosome. And, Lo includes actual experimental data establishing that his method works. *See* Ex. 1010 ¶¶ 95-96, 99-104; *id.* at FIGS. 3A, 4B, 5. A POSA would have no trouble applying this technique to multiple regions of a chromosome. A POSA would therefore find claim 2 obvious in view of Lo and the knowledge of the skilled artisan.

3. Claim 3: The method of claim 1, further comprising comparing a count from counts obtained in step (f) for a location on the reference genome to a count for the same location on the reference genome obtained from a reference sample, thereby estimating a relative copy number at the location

204. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 3.

205. The basic approach of Lo matches claim 3. Briefly, in Lo's method, one counts the number of sequence reads from a region of interest and computes a ratio of this number to the number of sequence reads from another comparator region or regions. In the simplest case, this ratio is a simple percentage of the reads from the chromosome of interest relative to the total number of sequence reads. *See supra* ¶¶ 191-192. This ratio is then compared to the ratio in another sample that is normal and does not show copy number change to see if there is a sequence read imbalance indicative of copy number change. *Id.* By doing this comparison, one compares the counts in step (f) to counts for the same location in a "reference sample," as the claim

requires. These are used to determine, for instance, relative copy number at the location. Disclosure of this approach is apparent throughout Lo.

206. Lo discloses, for instance, that “[i]n step 150, a parameter is determined from the first amount and the second amount. The parameter may be, for example, a simple ratio of the first amount to the second amount, or the first amount to the second amount plus the first amount...In step 150, the parameter is compared to one or more cutoff values.” *Id.* ¶¶ 62-64. The “cutoff value” of Lo is an empirically determined “numerical value whose value is used to arbitrate between two or more states (e.g. diseased and non-diseased) of classification for a biological sample. For example, if a parameter is greater than the cutoff value, a first classification of the quantitative data is made (e.g. diseased state); or if the parameter is less than the cutoff value, a different classification of the quantitative data is made (e.g. non-diseased state).” *Id.* ¶ 44.

207. As to the specific identity of the “cutoff,” Lo provides several disclosures, most often referring to values obtained from known normal samples. For instance, Lo claims “wherein at least one of the cutoff values represents a reference value established from one or more normal biological samples.” *Id.* at Claim 5.

208. Lo also provides a specific embodiment used for distinguishing aneuploid from euploid pregnancies, wherein the “parameter (e.g. the fractional

representation of the clinically relevant chromosome) is then compared to a reference range established in pregnancies involving normal (i.e. euploid) fetuses.”
Id. ¶ 65.

209. Such disclosure in Lo of using a cutoff from normal euploid pregnancies would be understood by a POSA to be a disclosure of using the claimed “reference sample” for determining relative copy number and, as Lo makes clear, the comparison to the reference sample is to the region of interest in the test same (i.e., the sequence reads in step (f)). This is made most clear by the disclosure in ¶¶ 99-101 in Lo, which discloses a formula for comparing the percentages of sequence reads from a chromosome of interest in the sample that is being tested to the percentage of reads from the chromosome of interest in a normal, control sample.
See id.

4. **Claim 6: The method of claim 1, further comprising (l) summing the counts from the counts obtained in step (f) for locations at a region of the reference genome; (m) summing the counts obtained from a reference sample for the same region of the reference genome; and comparing the sums of the counts obtained in step (l) and step (m), thereby estimating a relative copy number at the region of the reference genome.**

210. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 6.

211. Specifically, Lo teaches an approach for determining the existence of chromosomal imbalance in a biological sample obtained from a pregnant female.

The invention uses a “parameter of an amount of a clinically-relevant chromosomal region in relation to other non-clinically-relevant chromosomal regions (background regions) within a biological sample.” *Id.* ¶ 14. The approach then uses cutoff values as described above in ¶¶ 87, 193 to determine whether a “change compared to a reference quantity exists (i.e., an imbalance), for example, with regards to the ratio of amounts of two chromosomal regions (or sets of regions).” *Id.* This process estimates relative copy number at the region of the genome.

212. Specifically, Lo describes determining the percentage of sequences reads in a region of interest from the sample and comparing this percentage to that observed in a normal sample. *See id.* ¶ 99. This approach entails “summing the counts from the counts obtained in step (f) for locations at a region of the reference genome; (m) summing the counts obtained from a reference sample for the same region of the reference genome; and comparing the sums of the counts obtained in step (l) and step (m), thereby estimating a relative copy number at the region of the reference genome.” This is because the percentage involves summing counts from regions of interest. Indeed, the percentage is nothing more than an expression of the count of sequences at a region normalized to a total. As set forth in ¶¶ 99-101 of Lo, this percentage from the sample is compared to a normal by virtue of taking the difference between the sample and normal percentage and assessing how wide of a

gap there is in terms the number of standard deviations in the distribution of percentages observed in normal samples. *Id.* ¶¶ 99-101.

213. Therefore, a POSA would find claim 6 obvious in view of Lo and the knowledge of the skilled artisan.

5. Claim 7: The method of claim 1, wherein the unique tagged nucleic acid molecules are subject to a hybrid capture method prior to PCR in step (c) or the amplified tagged nucleic acid molecules produced in step (c) are subject to a hybrid capture method prior to step (d).

214. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 7.

215. Lo teaches an embodiment that uses hybrid capture prior to sequencing, wherein “the fraction of the nucleic acid pool that is sequenced in a run is further sub-selected prior to sequencing. For example, hybridization-based techniques such as oligonucleotide array could be used to first sub-select for nucleic acid sequences from certain chromosomes, e.g. a potentially aneuploid chromosome and other chromosome(s) not involved in the aneuploidy tested.” *Id.* ¶ 72.

216. Moreover, Lo establishes why a POSA would be motivated to do this and that there would be a reasonable expectation of success. As Lo explains, “[a]s mentioned above and established in the example section below, only a portion of the human genome needs to be sequenced to differentiate trisomy 21 from euploid cases. Thus, it would be possible and cost-effective to enrich the pool of nucleic acids to

be sequenced prior to random sequencing of a fraction of the enriched pool.” *Id.* ¶ 77; *id.* ¶ 79 (“The enriched pools of nucleic acids would then be subjected to random sequencing. This would allow the reduction in sequencing costs.”). Thus, as Lo makes clear, a POSA would be motivated to perform advance enrichment via hybrid capture because this would reduce the cost of the sequencing. As s Lo also teaches, there would be a reasonable expectation of success given a POSA’s knowledge of the widely available existing oligonucleotide array technologies.

217. Therefore, a POSA would find claim 7 obvious in view of Lo and the knowledge of the skilled artisan.”

6. Claim 8: The method of claim 7, wherein each of the unique tagged nucleic acid molecules subjected to the hybrid capture method has a nucleic acid sequence which is complementary to a capture probe on an array or on a microarray.

218. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 7. Lo further teaches the features of claim 8.

219. In my opinion, a POSA would understand that the hybrid capture method described in claim 7 works by isolating and enriching specific target DNA sequences of a sample through binding with probes that are complementary to the sequences of interest. Further, as noted above in connection with claim 7, Lo explicitly contemplates that this be formed through “hybridization based techniques

such as oligonucleotide array.” *See id.* ¶ 72. This is hybrid capture that uses an “array” or “microarray” as recited in the claim.

220. Therefore, a POSA would find claim 8 obvious in view of Lo and the knowledge of the skilled artisan.

7. Claim 11: The method of claim 1, wherein each of the tagged nucleic acid molecules comprise more than one tag.

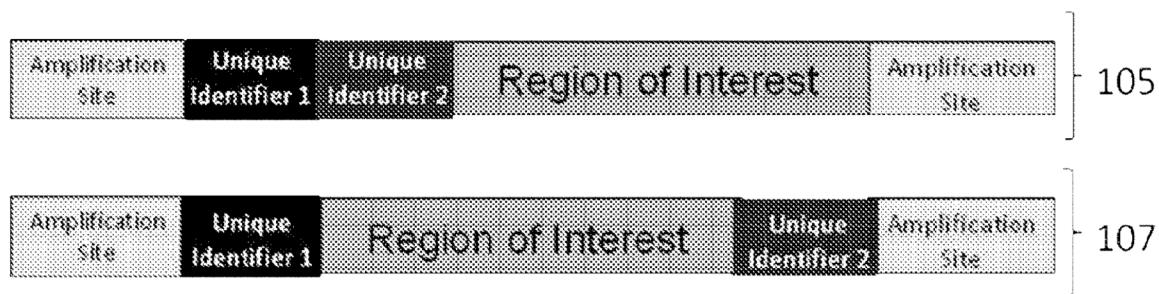
221. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 11, which recites “the method of claim 1, wherein each of the tagged nucleic acid molecules comprise more than one tag” would have likewise been rendered obvious by specific knowledge of the skilled artisan, as reflected in disclosure in each of McCloskey, Chee, Brenner, and Porreca. Below, I discuss the relevant disclosure in each of these references and explain why this claim would have been obvious.

222. McCloskey teaches the use of an “encoding nucleotide” containing multiple types of tags that are included in 4 distinct “informational regions” attached to each sequence of interest: “(1) the batch-stamp, unique to each experiment, specifies the DNA source such as the patient or sample identification, and the date of reaction; (2) the random barcode distinguishes among sequences arising from different cell or allele copies.” Ex. 1008 at 762. *See also id.* (“Each encoding oligonucleotide is unique because of distinct batch-stamp and random barcode

regions.”). Thus, McCloskey teaches tagged nucleic acid molecules that comprise more than one tag.

223. Chee also discloses the use of multiple tags or “identifiers” used as “constructs” for each Region of Interest (ROI) or sequence. As Chee explains, the “number of nucleotides used in the identifier will be determined in part by the number of different constructs that are to be produced in a set, but may range from a few nucleotides to up to 10, 20 or even 30 or more nucleotides in different sequence arrangements. In some cases, it is useful to use an additional unique identifier... The second unique identifier can either be found on the same side of the ROI... or flanking the ROI.... The use of two or even more unique identifiers can allow the use of shorter individual identifiers and/or two or more smaller ‘subsets’ of constructs due to the uniqueness conferred by the combination of the two tags.” Ex. 1015 at 20:4-15.

224. Diagrams labelled 105 and 107 depicted below in FIG. 1 of Chee illustrate the use of multiple tags on an individual sequence or ROI. Diagram 105 shows the “Unique Identifiers” or tags placed on the same side of the ROI, while diagram 107 places the Unique Identifiers on opposite sides of the ROI. *Id.*



225. Turning now to Brenner, it describes the use “indexing.” Specifically, Brenner describes a system of oligonucleotide tagging wherein regions of the oligonucleotide tag “can be allocated to identifying a restriction fragment or target that its associated selected probe is specific for. Such ‘indexing’ is useful when more than one target polynucleotide is being enumerated. For example, as in FISH assays to determine the extent that the ErbB2 gene is amplified, one selected probe can target a region of the ErbB2 gene, while another selected probe can target a region of the genome not expected to undergo any amplification in a patient. Such indexing can also be used to analyze multiple samples using high throughput DNA sequencing instruments, where tags from different patients have different indexing sequences.” Ex. 1014 at 21:20-32. Brenner’s teaching of using “indexing” to track the patient from which a DNA fragment originated in connection with high-throughput sequencing would be understood by the skilled artisan as a disclosure of using multiple tags (*i.e.*, one tag distinguishes molecules and is used for counting, and a second tag is the “index” that tracks the sample source).

226. Porreca, like each of Brenner, Chee, and McCloskey, also teaches multiple tags. Specifically, Porreca includes a disclosure much like Brenner and McCloskey's disclosure of using barcodes to track sample origin. Porreca even cites McCloskey for this purpose. As Porreca states, "barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 A1 (McCloskey et al.)." Ex. 1017 at 31:5-9. Just as with Brenner, Porreca's teaching of using a "barcode" to identify the sample from which a DNA fragment originated would be understood by the skilled artisan as a disclosure of using multiple tags (*i.e.*, one tag distinguishes molecules and is used for counting, and a second tag is the "index" that tracks the sample source).

227. A POSA would be motivated to utilize tagged nucleic acid molecules with more than one tag for several reasons. For instance, Lo provides a method for counting fragments of nucleic acid molecules by sequencing for determination of copy number information among multiple samples. A POSA would be motivated to use multiple tags so that one can not only distinguish fragments prior to amplification, but also to distinguish DNA arising from different samples. This would enable one to combine multiple samples into a single run on the sequencing instrument, which allows one to perform experiments more cheaply and quickly. As another advantage, the use of multiple tags allows one to effectively increase the

number of tags by enabling the possibility of combining individual tags in combinations, as in Chee, and this fact would have been well known to a skilled artisan at the time. This allows one to better ensure that sample molecules receive tags that can truly distinguish among identical molecules and improve the counting accuracy.

228. Further, for the same reason as I explain above in connection with claim 1, there can be no question of a reasonable expectation of being able to successfully use multiple tags because the basic machinery for tagging was well known and had been documented in many publications. *See supra* §§ XI.B-E, ¶¶ 144-164.

229. In the light of the above references, it would have been obvious for a POSA to incorporate more than one tag into a set of tagged nucleic acid molecules for use with Lo's method. .

8. Claim 17: The method of claim 1, wherein the genomic nucleic acids are obtained from blood or other bodily fluids.

230. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 17.

231. For example, Lo teaches an experiment wherein “[t]wenty milliliters of peripheral venous blood was collected from each subject. Maternal plasma was harvested after centrifugation at 1600xg for 10 minutes and further centrifuged at 16000xg for 10 minutes. DNA was then extracted from 5-10 mL of each plasma sample.” Ex. 1010 ¶ 92. Further, Lo discloses using plasma, urine, or serum as a

sample from which DNA is obtained. *Id.* ¶¶ 13-14, 106. Indeed, the examples and embodiments in Lo are generally based on the analysis of samples from blood or other bodily fluids.

232. In light of the above, Lo discloses and renders obvious this claim in my opinion.

9. Claim 18: The method of claim 1, wherein the genomic nucleic acids are obtained from tumor cell DNA in a bodily fluid.

233. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Lo further teaches the features of claim 18.

234. Lo discloses an embodiment of a method wherein “the cutoff value is determined at least in part on the percentage of tumor sequences in a biological sample, such as plasma, serum, saliva or urine, which contains a background of nucleic acid sequences derived from the non-malignant cells within the body.” *Id.* ¶ 52. Disclosing the analysis of tumor sequences in samples such as “plasma, serum, saliva or urine,” Lo discloses the analysis of genomic nucleic acids from a tumor cell DNA in a bodily fluid.

235. Therefore, Lo in view of the knowledge of the skilled artisan discloses and renders obvious this claim.

B. Ground 2: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In Brenner Or McCloskey Renders Obvious Claim 9

1. Claim 9: The method of claim 1, wherein each of the tagged nucleic acid molecules differs at more than one nucleotide.

236. A POSA would understand that Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 9, which recites “the method of claim 1, wherein each of the tagged nucleic acid molecules differs at more than one nucleotide,” is rendered obvious in further view of specific disclosure in Brenner or McCloskey.

237. I start with Brenner, which as I explain above, teaches tagging molecules for counting. Brenner teaches that the oligonucleotide tags should be selected “to be maximally different” so that they can be “distinguished even in the presence of sequencing errors.” Ex. 1014 at 21:14-18; *see also id.* at 25:57-58 (referring to “selection of tag sequences that are, in some sense, maximally different”); *id.* at 37:48-52 (“In another aspect, oligonucleotide tags are members of a mutually discriminable set; that is, oligonucleotide tags of such a set differ maximally from one another by some characteristic, such as sequence, melting temperature, or the like.”). Calling for the tags to be “maximally different,” Brenner unambiguously discloses tagged nucleic acid molecules that differ at more than one nucleotide. Indeed, the claimed difference of merely one nucleotide is the minimum

possible difference between nucleic acid sequences, and Brenner calls for the maximum possible difference.

238. Consider next McCloskey. McCloskey teaches that “[e]ach encoding oligonucleotide is unique because of distinct batch-stamp and random barcode regions.” Ex. 1008 at 762. Calling for each tag to be “unique” due to the combination of “distinct” batch-stamp and random barcode, McCloskey contemplates nucleic acid molecules that are different by more than one nucleotide, particularly because the possibility of the combined batch stamp and random barcode being identical save for a one-nucleotide difference is exceedingly slim. Further, McCloskey specifically provides guidance as to how many different barcodes can be generated based on the use of n nucleotides, thus enabling someone to increase tag diversity as needed. As McCloskey explains, the “number of distinguishable barcodes in a population of oligonucleotides used in a reaction is determined by the number of random bases, n . This enables one to distinguish among 4^n allele copies per reaction.” *Id.* at 764.

239. A POSA would have been motivated to use tagged nucleic acid molecules that differ by more than one nucleotide because by making the tags differ by more than a single nucleotide, one is more likely to ensure that the tags can be distinguished from one another despite potential sequencing errors. Brenner, for instance, teaches precisely this, stating that “[p]referably, sequences of

oligonucleotide tags are selected so that they can be distinguished even in the presence of sequencing errors. Thus, oligonucleotide tags (4120) are selected to be maximally different.” Ex. 1014 at 21:14-18.

240. Further, a POSA would have easily had a reasonable expectation of success in making the tagged nucleic acid molecules differ by more than one nucleotide. Indeed, Brenner teaches that there were well known ways of making the tags “maximally different.” Brenner explains, for instance, that there were “distance measures” known in the art for designing such sequences. Brenner cites several literature references to this effect, stating “exemplary distance measures for sequences applicable to the invention are well known and are widely disclosed in the literature, as exemplified by the following references: Gusfield, *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology* (Cambridge University Press, 1997); Navarro et al, *Flexible Pattern Matching on Strings* (Cambridge University Press, 2002); Sankoff et al, editors, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (Center for the Study of Languages, 1999); and the like. *Id.* at 25:58-66.

241. Claim 9 is thus obvious in view of Lo and the knowledge of the skilled artisan, in further view of specific disclosure in Brenner and McCloskey.

C. Ground 3: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In McCloskey, Brenner, or Porreca Renders Obvious Claims 12 And 14-16

1. Claim 12: The method of claim 1, wherein the tags further comprise a sample tag.

242. A POSA would understand that Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 12, which recites “wherein the tags further comprise a sample tag” is rendered obvious in further view of specific disclosure in Brenner, McCloskey, or Porreca.

243. At the priority date, “sample tags” were well-known and used regularly. Briefly, prior to sequencing, the DNA from each sample is given a tag that identifies that DNA as arising from a particular sample. Each sample has its own unique tag. The DNA from multiple samples can then be pooled and the pool of DNA can be sequenced at the same time. When doing the sequencing, one generates not just the sequence of the sample molecule, but also the tag that is attached to the molecule. Following sequencing, one can figure out which DNA sequence corresponded to a particular sample based on the tag sequence. Such sample tags were commercially available for widely used sequencing platforms as of at least 2009. *See, e.g.*, Ex. 1039.002 (“The MID Adaptors include a IO-nucleotide sequence tag on Adaptor A which is unique for each MID. When different libraries are prepared with different MIDs, they can be amplified by emPCR and sequenced together, in a multiplex

fashion; the sequencing reads can be deconvoluted by the data analysis software after the sequencing Run, such that the reads from each of the pooled libraries are identified by their MID tag and correctly assigned.”).

244. Each of Brenner, McCloskey, and Porreca disclose the use of a “sample tag” in this manner.

245. I start with Brenner. In discussing various options for formulating oligonucleotide tags, Brenner discloses the use of “indexing.” As Brenner explains, “[s]uch indexing can also be used to analyze multiple samples using high throughput DNA sequencing instruments, where tags from different patients have different indexing sequences.” Ex. 1014 at 21:29-32. A POSA would understand that this is a disclosure of the standard technique of using a tag on each molecule from a sample to label it as coming from that particular sample, sequencing the samples as a pool, and then deconvoluting the sequence data using the sample tag.

246. McCloskey similarly teaches a “batch stamp” tag is used alongside the “barcode” to identify specific samples. As McCloskey explains “[w]e have used batch-stamps of different lengths in different experiments. We typically use between six and eight base pairs to encode sample identity and date of reaction.. Each different nucleotide sequence yields a unique batch-stamp. The assigned batch stamp is registered for a specific sample run on a specific date and is used only once.” Ex. 1008 at 764; *see also id.* at 761 (“Batch-stamps specify the date and sample

identity, and barcodes detect template redundancy.”); *id.* at 763 (“the batch-stamp, unique to each experiment, specifies the DNA source such as the patient or sample identification....”). A POSA would understand that this is a disclosure of the standard technique of using a tag on each molecule from a sample to label it as coming from that particular sample, sequencing the samples as a pool, and then deconvoluting the sequence data using the sample tag.

247. Finally, Porreca also discloses the use of sample tags, specifically pointing to McCloskey for this purpose. Porreca teaches that “[i]n some embodiments, one or more of the techniques described herein may be combined in a single assay protocol for evaluating multiple patient samples in parallel.” Ex. 1017 at 21:3-5. To this end, Porreca explains that “barcodes may be included, for example, to uniquely identify individual samples (e.g., patient samples) that may be mixed together. (See, e.g., USPTO Publication Number US 2007/0020640 A1 (McCloskey et al.).” *Id.* at 31:5-9. This, again, is an unambiguous disclosure of the use of a “sample tag.” A POSA would understand that this is a disclosure of the standard technique of using a tag on each molecule from a sample to label it as coming from that particular sample, sequencing the samples as a pool, and then deconvoluting the sequence data using the sample tag.

248. A POSA would be motivated to utilize sample tags with Lo’s method. Lo’s approach seeks to identify fetal chromosomal defects in samples of maternal

blood through nucleic acid amplification, sequencing and subsequent counting of fragments, and also describes methods for comparing sequence counts across multiple samples. Providing a means to distinguish samples from each other through the use of tagging would permit someone to more efficiently analyze multiple samples at the same time on the sequencer. As reflected in Brenner, McCloskey, and Porreca the use of sample tags in this manner was a standard technique at the time and, accordingly, a POSA would have had a reasonable expectation of successfully using sample tags with Lo's method.

249. Therefore, a POSA would understand that Lo in light of McCloskey, Brenner or Porreca would render obvious the use of "tags that further comprise sample tags."

2. **Claim 14: The method of claim 1, wherein the unique tagged nucleic acid molecules are pooled with another plurality of unique tagged nucleic acid molecules having a different sample tag prior to the PCR or prior to said sequencing of the amplified tagged nucleic acid molecules of step (c).**

250. A POSA would understand that Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 14, which recites "wherein the unique tagged nucleic acid molecules are pooled with another plurality of unique tagged nucleic acid molecules having a different sample tag prior to the PCR or prior to said sequencing of the amplified tagged nucleic acid molecules of step (c)" is rendered obvious in further view of specific disclosure in Brenner, McCloskey, or Porreca.

251. Claim 14 recites nothing more than the standard manner of using a “sample tag” with sequencing, which I explain above in connection with claim 12. That is, the purpose of including a sample tag is to allow pooling of fragments from multiple samples prior to sequencing. Thus, for the same reasons that claim 12 would have been obvious, so too would claim 14 have been obvious.

3. Claim 15: The method of claim 14, further comprising deconvoluting the tag associated sequence reads by grouping the tag associated sequence reads.

252. As explained above, claim 14 would have been obvious to a POSA. Claim 15, which recites “deconvoluting the tag associated sequence reads by grouping the tag associated sequence reads” is rendered obvious in further view of specific disclosure in Brenner, McCloskey, or Porreca.

253. Claim 15 recites nothing more than the standard manner of using a “sample tag” with sequencing, which I explain above in connection with claim 12. That is, the purpose of including a sample tag is to allow for pooling of samples prior to sequencing and by necessity the tag associated sequence reads must be deconvoluted after sequencing so as to group reads accurately within each sample. Thus, for the same reasons that claims 12 and 14 would have been obvious, so too would claim 15 have been obvious.

4. Claim 16: The method of claim 1, further comprising obtaining copy number information from genomic nucleic acids of different species in a population using the method of claim 1, and determining a relative count of the genomic

nucleic acids of the different species in a population by comparing the copy number information obtained from the genomic nucleic acids of the different species.

254. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 16 is further rendered obvious in view of the disclosure of sample tagging in Brenner, McCloskey, and Porreca, which I address above in connection with claims 12 and 14-15.

255. Briefly, in a population that includes different species, one may wish to determine the relative genomic copy number of multiple species in the population and compare the copy numbers in those species. A skilled artisan would have been motivated to perform this task and would have had a reasonable expectation of successfully doing so given the skilled artisan's knowledge regarding sample tagging, as reflected by the disclosure of Brenner, McCloskey, and Porreca.

256. Specifically, given a POSA's knowledge of sample tagging, a POSA confronted with a multi-species population would have understood that he or she could easily analyze multiple species in the population at the same time using the available sequencing technology by using sample tagging. This would have made the process more efficient and less costly, and a POSA would have had a reasonable expectation of successfully using tags for this purpose because it was a standard approach that had been widely used. In short, claim 16 is obvious for the same reason that claims 12 and 14-15 are obvious.

D. Ground 4: The Combination Of Lo And The Knowledge Of A Skilled Artisan In Further View Of Specific Disclosure In Chee or Porreca Renders Obvious Claim 13

- 1. Claim 13: The method of claim 1, wherein the tagged nucleic acid molecules generated from step (b) and that differ by a single nucleotide are ignored during or after the counting step.**

257. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. It is my opinion that claim 13, which recites “the method of claim 1, wherein the tagged nucleic acid molecules generated from step (b) and that differ by a single nucleotide are ignored during or after the counting step” is obvious in further view of specific disclosure in Chee or Porreca.

258. Claim 13 generally pertains to the concept of insisting upon a high degree of consensus in deciding which sequence reads to count. As the '589 patent explains, “[t]hrough sequence error during processing, the tag may very occasionally mutate, creating the appearance of a new tagged nucleic acid molecule. But if the set of tags are sufficiently large, the tagged nucleic acid molecules will generally differ at more than one nucleotide, and so tagged nucleic acid molecules that differ by a single nucleotide can (optionally) be ignored when counting.” Ex. 1001 at 23:19-25.

259. In other words, when there is sequencing error, two sequences that ought to be the same can nonetheless appear different. The '589 patent contemplates eliminating these sequences that are out of the ordinary relative to other sequences.

This concept of generating a consensus sequence in this manner was previously taught in both Porreca and Chee.

260. Porreca teaches that specificity in detecting mutant allele may be increased by ignoring small differences between sequenced molecules. Porreca notes that “algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by *decreasing the minimum number of observed mutant reads necessary to make a consensus base-call*. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.” Ex. 1017 at 65:62-66:2. Thus, Porreca broadly teaches that one can tune the level of stringency in deciding whether to keep a sequence depending on how many reads must match each other.

261. Porreca also discloses in Example 7 a protocol that “permits detection of any of 50 specific known point mutations” in which reads “from each ‘sample’ are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.” *Id.* at 76:66-67; 78:1-8. Again, Porreca is disclosing tuning

the stringency with which one decides whether to keep certain sequence reads based on how they differ from other reads.

262. Turning to Chee, it discloses that some of the molecules of the method will contain “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 master set and can identify correct ROIs from ROIs that contain errors due to technical realities of chemical synthesis, amplification, and the like. In addition, the methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier.” Ex. 1015 at 10:33-46.

263. Chee goes on to explain that “[m]ultiple 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.” *Id.* at 14:43-50.

264. Thus, Chee, like Porreca, is disclosing tuning the stringency with which one decides whether to keep certain sequence reads based on how they differ from other reads. To the extent one identifies sequence reads that differ from other reads but that ought to be the same, one can infer that there were sequencing errors associated with those reads. Because the reads are potentially erroneous, one can ignore them.

265. A POSA of skill in the art would have been motivated to apply the methods of Chee and Porreca with tagging and ignore certain reads that differ from one another by a single nucleotide because it would improve the accuracy of the counting process. Specifically, where sequence information that ought to be the same is nonetheless different, it indicates that a sequencing error may have occurred and calls into question the veracity of the read. Eliminating reads that differ by a single nucleotide is a stringent standard that would allow one to only include reads that are most likely to be true and correct, a process that would improve the overall accuracy of the counting process. Those of skill in the art would have had no trouble implementing such a procedure because it amounts to nothing more than comparing sequence reads and eliminating reads that differ from what is expected.

Accordingly, there would have been a reasonable expectation of successfully implementing Chee and Porreca's error correction method with Lo.

266. Claim 13 would have thus been obvious to a POSA.

E. Ground 5: Claims 4 and 5 Are Unpatentable as Obvious Over Quake in View Of McCloskey, Brenner, Chee, or Porreca and Further in View of the Knowledge of a POSA

1. **Claim 4: The method of claim 1, further comprising g) summing counts obtained in step (f) for locations on a first region of the reference genome, wherein the first region comprises more than one location; h) summing counts obtained in step (f) for locations on a second region of the reference genome, wherein the second region comprises a number of locations which is comparable to the number of locations of the first region; i) comparing the counts obtained in step (g) to the counts obtained in step (h), thereby estimating a relative copy number at the first region of the reference genome to the copy number at the second region of the reference genome.**

267. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 4 would have been obvious in view of further disclosure in Quake.

268. As I explain above, the embodiments and examples in the Lo reference focus primarily on approaches where one assesses whether there is a copy number change by comparison to information from normal samples where one knows there is no copy number change. *See supra* ¶¶ 87-88, 191-193. Claim 4 merely recites the use of an internal standard, rather than an external standard, to assess copy number information. More specifically, it contemplates comparing sequence count

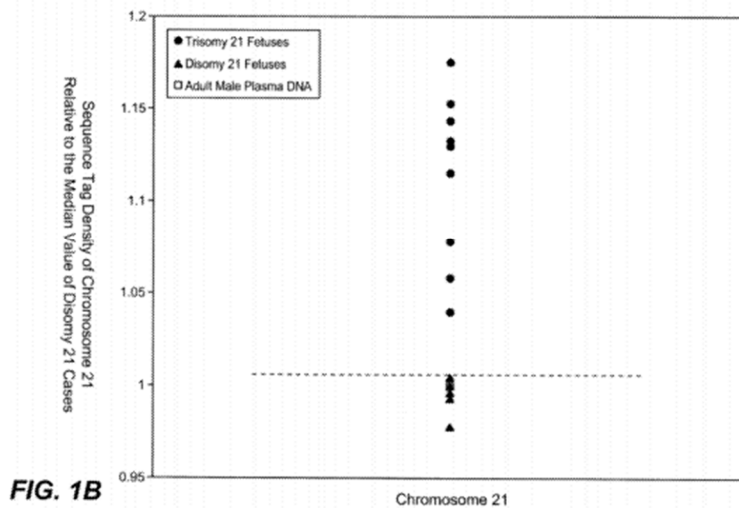
information from multiple locations of a genome region to sequence count information from locations of another genome region in the same sample and assessing copy number information by checking whether the first set of information is markedly different relative to the second set of information. As I explain above and further below, Quake teaches precisely this concept.

269. More specifically, Quake divides the genome into “windows” or “bins.” These windows are most often 50 kb in length, but Quake contemplates other bin sizes as well. *See supra* § XI.F.2. Then, one computes the median number of sequence reads mapping to all the bins across the chromosomes. *Id.* In this regard, Quake collects the sequence read count information at “locations on a first region” in a genome (*e.g.*, the bins on the chromosomes that one is interested in testing) and “locations on a second region in a genome” (*e.g.*, the bins on the other chromosomes that are not being tested and are presumed normal).

270. As Quake explains, by “taking the median M of all 22 autosomes, one obtains a normalization constant N that can be used to correct for differences in sequences obtained in different runs, as can be seen in Table 1. Thus, the normalized sequence tag density for chromosome 1 would be M_1/N ; for chromosome 22 it would be M_{22}/N .” Ex. 1019 at 7:44-54. Similarly, Quake explains, the “value of the sequence tag density is normalized within a sample. This can be done by counting the number of tags falling within each window on a chromosome; obtaining a

median value of the total sequence tag count for each chromosome; obtaining a median value of all of the autosomal values; and using this value as a normalization constant to account for the differences in total number of sequence tags obtained for different samples. *A sequence tag density as calculated in this way would ideally be about 1 for a disomic chromosome.*” *Id.* at 8:61-9:4.

271. Thus, rather than using an external control, Quake looks at large numbers of “locations” (i.e. 50 kb bins) across a “region” (a chromosome or set of chromosomes) in the genome in an individual sample, most of which will not include a copy number difference of any sort and will be normal, and then computes normalized densities. A “region” of the genome (e.g., a chromosome) with “locations” (i.e. 50 kb bins) with density that deviates from the normal sequence tag density of “about 1” can be identified as abnormal. This is shown, for instance, in Figure 1B, which shows that in samples where there is an extra copy of chromosome 21 the normalized sequence tag density is substantially above the ideal normal of 1:



Id. at FIG. 1; *see also id.* at Abstract (“The median count of autosomal values is used as a normalization constant to account for differences in total number of sequence tags is used for comparison between samples and between chromosomes.”).

272. In this regard, Quake compares the counts in the “locations on a first” region (*e.g.*, the bins on the chromosomes that one is interested in testing) to the counts in the “locations on a second region” (*e.g.*, the bins on the chromosomes are not being tested and are presumed normalized). The normalized counts in the latter should ideally be “about 1,” while the former will be higher if there is a copy number difference. Quake uses this comparison to provide relative copy number information at the first location compared to the second. Specifically, Quake’s method establishes that the first region has an extra copy of the chromosome in question.

273. Quake provides a further summary of his method below:

Using a reference sequence, one assigns the sequence tags to their corresponding chromosomes including at least the specified

chromosome by comparing the sequence to reference genomic sequence. Often there will be on the order of millions of short sequence tags that are assigned to certain chromosomes, and, importantly, certain positions along the chromosomes. ***One then may determine a first number of sequence tags mapped to at least one normally distributed chromosome portion and a second number of sequence tags mapped to the specified chromosome portion, both chromosomes being in one mixed sample.*** The present method also involves correcting for nonuniform distribution sequence tags to different chromosomal portions. This is explained in detail below, where a number of windows of defined length are created along a chromosome, the windows being on the order of kilobases in length, whereby a number of sequence tags will fall into many of the windows and the windows covering each entire chromosome in question, with exceptions for non-informative regions, e.g., centromere regions and repetitive regions. Various average numbers, i.e., median values, are calculated for different windows and compared. By counting sequence tags within a series of predefined windows of equal lengths along different chromosomes, more robust and statistically significant results may be obtained. ***The present method also involves calculating a differential between the first number and the second number which is determinative of whether or not the abnormal distribution exists.***

Id. at 4:40-67. The key language here is that one computes counts in first regions to generate a “first number,” and counts in second regions to generate a “second

number,” and a “differential” is calculated to determine whether a copy number change is present. This corresponds exactly to claim 4.

274. As Quake explains, his method does not require the use of an external standard and instead uses an internal standard. *See id.* at 9:6-10 (“This method does not require the use of an external standard, but, rather, provides an internal reference, derived from al [sic] of the sequence tags (genomic sequences), which may be, for example, a single chromosome or a calculated value from all autosomes.”).

275. To the extent Quake’s examples generally involved computing a median and not a sum of the counts of locations of first and second region, this is a minor distinction and POSA would have found it obvious to, for instance, compute an alternative identifier of central tendency across the bins, such as an average, which would involve computing a sum across the bins (*i.e.*, “locations”) across the chromosomes (*i.e.*, “regions”). As Quake explains, for instance, “[v]arious average numbers, *i.e.*, median values, are calculated for different windows and compared.” *Id.* at 4:59-61. Further, Quake explains that means can be used in place of medians assuming one removes outlier. *See id.* at 29:24-26 (“We do not expect the confidence interval of the median and the mean to be very different if the small number of outliers has been removed.”). Indeed, there are only a small number of indicators of central tendency of the values in a data set, with the average and median

being by far the most common. A POSA would have thus found it obvious to use the average instead of a median.

276. To be clear, Quake's method is not limited to entire chromosomes. In fact, Quake teaches that to increase sensitivity its method can focus on certain regions of the genome by selecting in advance DNA sequences from regions of interest. As Quake states, "[a]nother method for increasing sensitivity to fetal DNA is to focus on certain regions within the human genome. One may use sequencing methods which select a priori sequences which map to the chromosomes of interest (as described here elsewhere, such as 18, 21, 13, X and Y). One may also choose to focus, using this method, on partial chromosomal deletions, such as 22q11 deletion syndrome. Other microdeletions and microduplications are set forth in Table 1 of US 2005/0181410, published Aug. 18, 2005 under the title 'Methods and apparatuses for achieving precision genetic diagnosis.'" *Id.* at 13:52-63.

277. A POSA would be motivated to use Quake's internal standard method in place of the Lo's method because it would obviate the need to use any sort of external samples as references. Instead, one can confidently determine from just a single sample whether there is a copy number change in a particular region of the genome by looking for outlier regions within a single genome. This would allow a POSA to utilize the sequencing instrument's multiplexing capacity for more samples that are being tested rather than control samples that are not being tested for copy

number changes. A POSA would have a reasonable expectation of successfully using this approach because Quake proved that it would work, including with actual experimental data in his patent filing, which I show above.

2. **Claim 5: The method of claim 4, wherein step (h) further comprises j) summing counts obtained in step (f) for locations at a third region of the reference genome, wherein the third region comprises a number of locations which is comparable to the number of locations of the first region; and k) obtaining an average of the sum of the counts obtained in step (f) for locations at the second region and the sum of the counts obtained in step (j) for locations at the third region.**

278. As discussed above, Quake in view of the knowledge of the skilled artisan renders obvious claim 4. Quake further teaches the features of claim 5.

279. Claim 5 adds little relative to claim 4. Claim 4 requires that one use two regions of a reference genome and claim 5 simply requires that one uses three regions and that one average the counts in each of these regions. Quake discloses this or, at a minimum, renders this obvious.

280. To begin Quake discloses step (j) of obtaining counts from a third region (*i.e.*, a chromosome). Indeed, the method in Quake obtained sequence tag counts not just for two chromosomes, but for all autosomes. *Id.* at 22:39-41 (“we normalized the sequence tag density of each chromosome (except chromosome Y) to the median sequence tag density among autosomes.”); *see also id.* at FIG. 12 (plotting the T-statistics for each chromosome in several patients). While referring

to generating a “median sequence tag density among autosomes,” as explained above in connection with claim 4, it would have nonetheless been obvious to alternatively compute an average, which would require computing a sum.

281. Quake describes his method in more details as follows. As Quake explains, “a sliding window of 50 kb was applied across each chromosome, except in regions of assembly gaps and microsatellites, and the number of sequence tags falling within each window was counted and the median value was chosen to be the representative of the chromosome. Because the total number of sequence tags for each sample was different, for each sample, we normalized the sequence tag density of each chromosome (except chromosome Y) to the median sequence tag density among autosomes. The normalized values were used for comparison among samples in subsequent analysis.” *Id.* at 22:32-43. Thus, Quake again describes determining median numbers of sequence counts across different autosomes, which satisfies steps (j) and (k) of this claim., or at a minimum renders obvious the determination of other well-known indicators of central tendency, such as an average (which requires a sum). *See id.* at 4:1-5 (“The median count of autosomal values (i.e., number of sequence tags per autosome) is used as a normalization constant to account for differences in total number of sequence tags...between samples and between chromosomes.”); *id.* at 23:32-37 (explaining in Example 7 that because “the total number of sequence tags obtained for each sample was different, in order to

compare among samples, we normalized each chromosomal sequence tag density value (except chromosome Y) by the median sequence tag density among all autosomes (non-sex chromosomes)"). Given humans have 22 autosomal chromosomes, the method in Quake describes determining copy number by comparing a target region to an average of at least two other regions in the reference genome.

282. A POSA would be motivated to use this approach and have a reasonable expectation of successfully doing so for the same reasons explained above in connection with claim 4.

F. Ground 6: Claims 10 is Unpatentable as Obvious Over Lo In View Of The Knowledge Of A POSA In Further View Of Specific Disclosure In Chee, Porreca, Or Brenner

- 1. Claim 10: The method of claim 1, wherein said tagging the segments with nucleic acid tags comprises ligation of adaptors comprising the tags to at least one end of the segments of the genomic nucleic acids.**

283. As discussed above, Lo in view of the knowledge of the skilled artisan renders obvious claim 1. Claim 10, which recites the use of ligation to introduce tags, would have also been obvious in further view of disclosure in Chee, Porreca, or Brenner.

284. A skilled artisan would know that tagging of nucleic acid molecular fragments can be accomplished using multiple tags in a small number of well-established ways. As of the earliest priority date of the '589 patent, it was well-

known and routine in the art to ligate adaptors comprising tags to at least one end of the segments of the genomic nucleic acids. Indeed, the specification of the '589 patent states as much:

Miner et al. (2004) describe a method of molecular barcoding to label template DNA prior to PCR amplification, and report that the method allows for the identification of contaminant and redundant sequences by counting only distinctly tagged sequences (22). U.S. Pat. No. 7,537,897 describes methods for molecular counting by labeling molecules of an input sample with unique oligonucleotide tags and subsequently amplifying and counting the number of different tags (23). Miner et al. and U.S. Pat. No. 7,537,897 both describe labeling of input nucleic acid molecules by ligation, which has been found to be an inefficient reaction.

Ex. 1001 at 2:12-22.

285. Because the use of ligation was a standard technique in the toolkit of the skilled artisan, it was disclosed in other references that describe tagging as well.

286. Chee, for instance, teaches that a “diverse but pre-defined set of sequences can be synthesized and attached to the ROIs, e.g., via PCR primers or by ligation” and that there are a “wide variety of methods that can be used to assemble oligos into larger constructs, and these can be used to assemble larger nucleic acids using the isolated ROIs obtained from the present methods.... [S]ome examples

include PCR with the oligo set; ligation followed by PCR; [and] ligation followed by cloning in a biological vector.” Ex. 1015 at 11:18-20 and 23:28-35.

287. Porreca likewise discloses the use of ligation to incorporate random tag differentiator sequences. “For example, nucleic acids comprising differentiator tag sequences may be incorporated by ligation. This is a flexible method, because molecules having differentiator tag sequence can be ligated to any blunt-ended nucleic acids....Alternatively, the sequencing adaptors can be synthesized with the random differentiator tag sequences at their 3' end (as degenerate bases), so that only one ligation must be performed.” Ex. 1017 at 31:10-21.

288. Additionally, Brenner discloses that one technique for providing polynucleotides with unique nucleotide tags is by ligation: “For example, polynucleotides of a population to be labeled are each associated or linked, e.g. by ligation, to an oligonucleotide tag from a population that has a much larger size than that of the target polynucleotide population.” Ex. 1014 at 17:60-64.

289. A POSA would have been motivated to use ligation to introduce tags for several reasons. First, there are only a small number of ways to introduce tags. Besides ligation, primer extension is the only other realistic method that one would use to introduce tags. Because there is only a small number of ways to introduce tags, a POSA would be immediately motivated to consider ligation as an option for doing this, putting aside any considerations as to its difficulty or effectiveness.

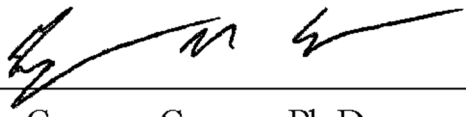
Second, upon recognizing that there were only a small number of ways to introduce tags, a POSA would have been particularly motivated to use ligation because it was such a well-known, routine, and conventional approach for introducing tags that was known to work well. This is why it is contemplated in each of Brenner, Porreca, and Chee.

290. Finally, because ligation was a conventional technique that was tried and true, a POSA would have had a reasonable expectation of successfully using ligation to introduce tags.

DECLARATION

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are belief 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.

Dated: July 25, 2025

By: 
Gregory Cooper, Ph.D.