Filed on behalf of: Sequenom, Inc.
By: Steven P. O'Connor
Michele C. Bosch
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, L.L.P.
901 New York Avenue, NW
Washington, DC 20001-4413
Telephone: 202-408-4000
Facsimile: 202-408-4400
E-mail: steven.oconnor@finnegan.com
michele.bosch@finnegan.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SEQUENOM, INC.
Petitioner

v.

THE BOARD OF TRUSTEES OF
THE LELAND STANFORD JUNIOR UNIVERSITY
Patent Owner

Patent 8,195,415

DECLARATION OF STACEY BOLK GABRIEL

SEQUENOM EXHIBIT 1010

Table of Contents

I, Stacey Bolk Gabriel, declare as follows:

I. Introduction

1. I have been retained by Sequenom, Inc. ("Petitioner") as an independent expert consultant in this proceeding before the United States Patent and Trademark Office. Although I am being compensated at my rate of $500 per hour for the time I spend on this matter, no part of my compensation is dependent on the outcome of this proceeding, and I have no other interest in this proceeding.

2. I understand that this proceeding involves U.S. Patent No. 8,195,415 ("the '415 patent") (Ex. 1001), the application for which was filed on January 29, 2010, as U.S. Patent Application No. 12/696,509, and issued on June 5, 2012. I also understand that the '415 patent is what is referred to as a "divisional" of U.S. Patent Application No. 12/560,708, which was filed on September 16, 2009, which in turn claims priority to Provisional Application No. 61/098,758, filed September 20, 2008. I further understand that the '415 patent indicates it is assigned to the Board of Trustees of the Leland Stanford Junior University ("Patent Owner").

3. I have been asked to consider whether a person of ordinary skill in the art would have understood that certain references teach, either alone or in combination, the features recited in the claims of the '415 patent. My opinions are set forth below.

II. Qualifications

4. I received a Bachelor of Sciences degree from Carnegie Mellon University in Molecular Biology in 1993. I received a Ph.D. in Genetics in 1998 from Case Western Reserve University. I conducted my thesis research projects under the direction of Dr. Aravinda Chakravarti using genomic mapping techniques and linkage analysis to identify genes involved in genetic diseases. My graduate research focused on characterizing genes involved in idiopathic congenital central hypoventilation syndrome, a rare disorder of respiratory control, and Hirschsprung (HSCR) disease, the most common cause of congenital intestinal obstruction.

5. My graduate research involved searching for sequence mutations in DNA by using techniques such as polymerase chain reaction (PCR), microsatellite genotyping, and DNA sequencing. I conducted genotyping on members from 61 families containing individuals with and without HSCR to study the inheritance pattern of the disease. I performed fluorescent dye-terminator cycle sequencing (based on the first generation Sanger dideoxy sequencing method) using PCR with genomic DNA in a primer extension sequencing reaction. The PCR products were run out (electrophoresed) on a slab gel and an automated ABI 377 DNA Sequencer was used for data collection. I then performed linkage analyses of the data by comparing DNA sequences from HSCR affected and non-affected individuals to

search for differences (polymorphisms) in the sequences. This study identified three important regions of the genome to explain the inheritance of HSCR (only one of these regions was previously known). It also showed that some of these mutations are in non-protein coding regions, suggesting the importance of noncoding variation. This experiment was an early example of complete genetic dissection of a multifactorial disorder.

6. From November 1998 to February 2002, I was a Research Scientist in the Functional Genomics Program of the Whitehead Institute Center for Genome Research, now referred to as the Medical and Population Genetics Program of the Broad Institute of Harvard and MIT ("Broad Institute"). My responsibilities included laboratory work involving technology development for Single Nucleotide Polymorphism (SNP) genotyping, supervising technicians, and creating assays for SNP genotyping. During that time, I worked on the technical development and implementation of the first genotyping platforms to be used at our institute for high throughput SNP genotyping. All of these platforms utilized the basic PCR technique or a variation of PCR at some step to amplify the individual pieces of DNA; however, each platform used a different strategy and method of detection. For example, I worked on TaqMan assays (assays that use allele specific fluorescent probes designed to increase the specificity of real-time PCR assays) and spotted array designs (hybridization techniques that use small fragments of

PCR products that correspond to mRNAs) to genotype SNPs. Specifically, I helped design a method for parallel genotyping of SNPs called single base extension-tag array on glass slides (SBE-TAGS). This method uses techniques such as multiplex PCR (amplification of genomic DNA using multiple primers), primer extension using fluorescently labeled dideoxynucleotide triphosphates (ddNTPs), and DNA spotted microarrays. The ScanArray 5000 (GSI Luminonics) was used to scan the fluorescent signal for genotyping. With this study we were able to genotype over 100 SNPs, obtaining over 5,000 genotypes with approximately 99% accuracy.

7. During my time as a Research Scientist in the Functional Genomics Program, I used the genotyping methods described above to investigate the haplotype structure of the human genome. I designed genotyping experiments in SNPs in 275 individuals from Africa, Europe, and Asia. Using multiplex PCR followed by primer extension, the DNA sample was loaded onto a microarray chip (SpectroCHIP, Sequenom) and analyzed by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) using a Broker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom). We characterized haplotype patterns across 51 autosomal regions (spanning 13 megabases of the human genome) using this method. This research resulted in a first author Science publication (Gabriel et al. Science 296(5576):2225-2229 (2002)), which is widely

regarded as laying the foundation for the International Human HapMap project. The International Human HapMap project is a multi-country collaboration to develop a haplotype map (Hap Map) of the human genome based on SNP genotyping. The data is publicly released by researchers from participating countries and is a key resource for researchers to find genetic variants affecting health, disease, and responses to drugs and environmental factors.

8. From February 2002 to May 2003, I was the Scientific Director of the SNP genotyping and Hap Map Program of the Whitehead Institute Center for Genome Research. As Scientific Director, I was responsible for all aspects of the Center's contribution to the International HapMap Project. At the Whitehead HapMap Program I oversaw a team of 15 technicians, analysts, and software engineers, played an active role in project design and quality control, and served on the International HapMap project Steering committee.

9. From May 2003 to May 2004, I was the Associate Director of the High Throughput Biology, Medical and Population Genetics Program of the Whitehead Institute Center for Genome Research. As Associate Director, I spearheaded the expansion of SNP genotyping activity from targeted activity for the Human HapMap project to a centralized technology platform with dedicated activity in technology development, large-scale production, data management, and

analysis. I also oversaw the successful completion of the Whitehead Institute's contribution to the Human Hap Map project, which had a $10 million budget.

10. From May 2004 to January 2009, I was the Director of the Genetic Analysis Platform of the Broad Institute. As Director, I was responsible for creating, scaling and directing the Genetic Analysis Platform of the Broad Institute. The Genetic Analysis Platform encompassed all production and data management activities related to nucleic acid analysis including gene expression, genotyping and re-sequencing. During the Platform's peak period from 2006 to 2008, I operated the platform with yearly revenues of $45 million, and oversaw a staff of 65 individuals including project managers, research scientists, software engineers, and computational biologists. One of the key milestones of the Genetic Analysis Platform included producing microarray data on over 100,000 DNA samples over an 18 month period. I also directed data production for over 50 publications describing genome-wide association findings. Massively parallel sequencing using micro arrays was used in many of these studies for SNP genotyping. DNA genomes of individuals with and without the disease of interest were compared to identify common variations in the genome that are associated with the disease. These studies focused on identifying genes involved in different diseases such as cancer, diabetes, arthritis, multiple sclerosis, and cardiovascular diseases. In contrast to other methods which specifically test one or a few genetic regions,

these genome-wide association studies investigated the entire genome of individuals.

11. From January 2009 to May 2012, I was Co-Director of the Genome Sequence and Analysis Program and Medical and Population Genetics Program of the Broad Institute. As the Co-Director, I was responsible for planning, execution, and delivery of a portfolio of cancer and medical sequencing projects as part of the National Human Genome Research Institute (NHGRI) large-scale sequencing grant. I was also a Co-Principal Investigator with Eric Lander for a large-scale sequencing grant renewal. As Co-Director and Principal Investigator, I secured over $100 million in other NIH awards over a period of 5 years aimed at large scale genotyping and sequencing. As Co-Director, I directed the activity of cross-disciplinary teams totaling 60 people, including project managers, analysts, computational biologists and software engineers in the analysis of massively parallel sequence data as applied to an array of cancer genomics and medical genetics projects. As Co-Director, I served as co-chair of the Data Production committee for the International 1000 Genomes Project, as well as serving as a member of the Executive and Steering committee for The Cancer Genome Atlas.

12. As Co-Director, I was involved in developing a technique called Solution Hybrid Selection (SHS), which is used to prepare specific regions of the genome for massively parallel sequencing using the Illumina platform. Because of

the large size of the human genome, it is more feasible in some cases to sequence only certain regions of the genome. The SHS technique uses RNA "baits" to "fish" pieces of DNA out of a "pond" of DNA fragments. PCR is used at two different stages to amplify the DNA. Additionally, quantitative PCR is used to quantify the final amount of DNA that was "caught" by the "bait." The resulting DNA was sequenced using the Illumina platform, but this technique can be used on any sequencing platform. This method has been commercialized by Agilent Technologies as "SureSelect" and is the leading product for genome selection today.

13. Since May 2012, I have been the Director of the Genomics Platform of the Broad Institute. As Director, I am in charge of the Broad Institute's largest platform, and the largest US genome center, comprising 180 people dedicated to all sample handling, microarray, genotyping, and sequencing activities. I am responsible for a $90 million annual budget for genomic activities. I oversee project management and data analysis activities, primarily in support of cancer, and medical genetics, as well as technology development and evaluation and implementation of new technology platforms. I also maintain all the leadership activities I described above as Co-Director of the Genome Sequence and Analysis Program and Medical and Population Genetics Program.

14. Throughout my research experience I have used a variety of genomic tools including PCR, genotyping (for example by single base extension, hybridization, or oligo ligation), and sequencing (for example by Sanger sequencing or massively parallel sequencing).

15. All of the genomic technologies use methods such as template preparation (preparation of pieces of DNA to be sequenced), sequencing and imaging, and data analysis. However, the unique combination of specific techniques used within these methods is what distinguishes one technology from another. I have had the opportunity to use and help develop numerous platforms that utilize very different techniques. I have participated in the development and use of multiple sequencing platforms, including both Sanger type sequencers and massively parallel DNA sequencers that utilize different strategies to sequence DNA.

16. I have served and continue to serve on various editorial and advisory boards related to genomic research. For example, from February 2007 to the present, I have served on the External Advisory Committee for National Heart, Lung, and Blood Institute (NHLBI) Resequencing and Genotyping Service. From July 2009 to June 2013, I was a standing member of the NIH Study Section of Genomics, Computational Biology and Technology. From May 2010 to the present, I have served on the Scientific Advisory Board of Genome Canada. I have

served on the editorial boards of Human Genetics and Genome Research. My additional peer review and other professional activities are set forth on my curriculum vitae, a copy of which is submitted herewith as Ex. 1011.

17. I have authored over 90 peer-reviewed publications. As my research has been primarily directed to genome sequencing, most of these publications involve the application of sequencing technology to the study of human disease. DNA sequences of individuals with and without a specific disease were compared in order to determine whether there is a common genetic variable in those individuals with the disease. These publications resulted in the identification of genes and mutations that are associated with diseases including cancer, diabetes, arthritis, multiple sclerosis, and cardiovascular diseases. Additionally, I have published protocols for methods that I have helped develop to prepare DNA for use in massively parallel sequencing.

18. I have presented lectures at a variety of academic and industry conferences, and lecture about 6 to 8 times a year at conferences involving genomics. For example, I have presented at conferences held by the International Congress of Human Genetics, the American Society of Human Genetics, the American Association for Cancer Research, the American Heart Association, the Multiple Myeloma Research Foundation, and the Association for Research in

Vision and Ophthalmology. These presentations were primarily focused on using genomics to understand the genetic basis of human disease.

19. I am not an attorney and offer no legal opinions. My curriculum vitae, which includes a more detailed summary of my background, experience, and publications, is attached as Ex. 1011.

III. Summary of Opinions

20. All of the opinions contained in this Declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming the opinions expressed in this Declaration, I reviewed the (1) '415 patent (Ex. 1001); (2) portions of the prosecution history for the '415 patent; (3) U.S. Patent Application Publication No. 2009/0029377 to Lo *et al.* ("*Lo II*") (Ex. 1002); (4) U.S. Provisional Patent Application No. 60/951,438 to Lo *et al.* ("*Lo I*") (Ex. 1003); (5) U.S. Patent Application Publication No. 2005/0221341 to Shimkets *et al.* ("*Shimkets*") (Ex. 1004); (6) Tian-Li Wang *et al.*, "Digital karyotyping," Proc. Natl. Acad. Sci. USA, 99(25):16156-61 ("*Wang*") (Ex. 1005); (7) LaDeana W. Hillier, "Whole-genome sequencing and variant discovery in *C. elegans*," Nature Methods, 5(2):183-88 (and on-line supplementary information) ("*Hillier*") (Ex. 1006); (8) Juliane C. Dohm *et al.*, "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing," Nucleic Acids Res., 36(16):e105 ("*Dohm*") (Ex. 1007); (9) U.S. Patent No. 7,888,017 to Quake and Fan ("*Quake*")

(Ex. 1008); and (10) Andrew D. Smith *et al.*, "Using quality scores and longer reads improves accuracy of Solexa read mapping," BMC Bioinformatics, 9:128 (*"Smith"*) (Ex. 1009), while drawing on my experience and knowledge of genomic sequencing and related molecular biology techniques.

21. My opinions have been also guided by my appreciation of how a person of ordinary skill in the art would have understood the claims of the '415 patent at the time of the alleged invention, which I have been asked to assume is September 20, 2008.

22. At the time of the alleged invention, a person of ordinary skill in the art relevant to the subject matter of claims 1 through 17 of the '415 patent would have a multi-disciplinary background. That person would have at least a bachelor's degree in a life sciences area (e.g., biology, cell biology, genetics, and molecular biology) and at least a master's degree or Ph.D. in computational biology, mathematics or statistics, or equivalent training. A person of ordinary skill in the art should understand both the operation and application of massively parallel DNA sequencing platforms, and have significant direct experience at performing and applying these techniques. Further, a person of ordinary skill in the art should understand and have experience with techniques for aligning sequence reads generated by massively parallel sequencing to a reference genome.

23. It is my understanding that a claim is anticipated by the prior art if a prior art reference discloses each and every feature of the claim. Also, I understand that when the prior art discloses a species that falls within a genus, or range, a claim to the genus, or range, is anticipated by that prior art species.

24. It is my understanding that a claim is unpatentable over the prior art if the differences between the features in the claim and the prior art are such that the subject matter of the claim as a whole would have been obvious at the time of the invention to a person having ordinary skill in the pertinent art. I understand that in some circumstances a teaching, suggestion, or motivation in the prior art would have led a person of ordinary skill in the art to modify a reference, or combine references, to arrive at the claimed invention. I also understand there may be other reasons why a claim would have been obvious. For example, I understand that it would be obvious for a person of ordinary skill in the art to use a known technique to improve a similar method in the same way and yield predictable results. I also understand it would be obvious for a person of ordinary skill in the art to combine prior art teachings to achieve a certain desired result with a reasonable expectation of success.

25. Based on my experience and expertise, it is my opinion that certain references teach, alone or in combination, all of the features recited in the claims of the '415 patent.

IV. Overview of the '415 Patent

26. I understand that the '415 patent is directed to "a method to achieve digital quantification of DNA (i.e., counting differences between identical sequences) using direct shotgun sequencing followed by mapping to the chromosome of origin and enumeration of fragments per chromosome." Ex. 1001, '415 patent, Abstract. "Shotgun sequencing" refers to random sequencing of nucleic acid fragments in a sample.

27. According to the '415 patent, "[t]here is therefore a desire to develop non-invasive genetic tests for fetal chromosomal abnormalities." *Id.*, 1:52-54. The '415 patent addresses that desire by providing methods for analyzing a maternal sample, such as blood, which contains maternal and fetal DNA, for detecting fetal aneuploidy. As explained in the '415 patent, "[t]he abnormal distribution of a fetal chromosome or portion of a chromosome (i.e., a gross deletion or insertion) may be determined in the present method by enumeration of sequence tags as mapped to different chromosomes." *Id.*, 3:64-4:1. The methods entail "carr[ying] out sequence determinations on the DNA fragments in the sample, obtaining sequences from multiple chromosome portions of the mixed sample to obtain a number of sequence tags of sufficient length of determined sequence to be assigned to a chromosome location within a genome [by comparison to a reference sequence] and of sufficient number to reflect abnormal distribution." *Id.*, 4:34-43.

28. The '415 patent applies conventional statistical data analysis techniques to the sequencing data obtained from the methods. For example, according to the '415 patent one may normalize the data obtained from the methods to provide more robust and statistically significant results. In one approach, non-uniform distribution of sequence tags to different chromosomal portions may be corrected by using windows of defined length to subdivide the chromosomes. *Id.*, 4:51-67. This same approach to data analysis can be used to correct for the known bias resulting from the G/C content of the maternal and fetal DNA sequenced in the methods claimed in the '415 patent. *Id.*, 5:23-30.

V. Claim Construction

29. I understand that in this type of proceeding before the United States Patent and Trademark Office, a claim receives the broadest reasonable interpretation in light of the specification of the patent in which it appears. I also understand that, at the same time, claim terms are given their ordinary and accustomed meaning as would be understood by a person of ordinary skill in the art. But I also understand that a patentee may act as his own lexicographer in redefining the meaning of particular claim terms away from their ordinary meaning. I have followed these principles in my analysis. I discuss a few terms below and what I understand to be Petitioner's constructions of these terms, which I agree with.

A. Chromosome Portion

30. Each of independent claims 1 and 13 recites testing for or determining a "chromosome portion." Ex. 1001, 33:53-34:58; 36:1-17. I understand that the Petitioner has offered the broadest reasonable construction of the term "chromosome portion" consistent with the specification as "either an entire chromosome or a significant fragment of a chromosome." I have used this construction in my analysis and agree with it because the '415 patent specifically defines the term this way. *See id.*, 4:5-7.

B. Window

31. Independent claim 1 recites determining values for a number of sequences tags using "a number of windows of defined length." Ex. 1001, 33:33-34:58. The '415 patent treats the terms "window" and "bin" as equivalent. Ex. 1001, 7:37. I understand that the Petitioner has offered the broadest reasonable construction of the term "window" or "bin" consistent with the specification as a "predefined subsection of a chromosome." I have used this construction in my analysis and agree with it because the specification of the '415 patent supports such an interpretation:

- "Each autosome (chr. 1-22) is computationally segmented into contiguous, non-overlapping windows" and "[e]ach window is of sufficient length to contain a significant number of reads (sequence

tags, having about 20-100 [bp] of sequence)...." Ex. 1001, 5:4-9.

- "The present method also involves correcting for nonuniform distribution [of] sequence tags to different chromosomal portions [using windows]." *Id.*, 4:51-52.
- "[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." *Id.*, 4:53-59.

C. Sliding Window

32. Independent claim 13 recites that each chromosomal portion comprises "a sliding window of a predetermined length." Ex. 1001, 36:1-17. I understand that the Petitioner has offered the broadest reasonable construction of the term "sliding window" consistent with the specification as "contiguous, overlapping or non-overlapping, predefined subsections of a chromosome." I have used this construction in my analysis and agree with it because the specification of the '415 patent supports such an interpretation:

- "Each autosome (chr. 1-22) is computationally segmented into contiguous, non-overlapping windows. (A sliding window could also

be used)." Ex. 1001, 5:4-6.

- "Because the distribution of sequence tags across each chromosome was non-uniform (possibly technical artifacts), we divided the length of each chromosome into non-overlapping sliding window[s] with a fixed width (in this particular analysis, a 50 kbp window was used), skipping regions of genome assembly gaps and regions with known microsatellite repeats." *Id.*, 23:14-20.

D. Sequence Tag Density

33. Claims 2 and 10-12 recite comparing or calculating a "sequence tag density." Ex. 1001, 34:59-64; 35:16-33. I understand that the Petitioner has offered the broadest reasonable construction of the term "sequence tag density" consistent with the specification as "the normalized value of sequence tags for a defined window of a sequence on a chromosome ... where the sequence tag density is used for comparing different samples and for subsequent analysis." I have used this construction in my analysis and agree with it because the '415 patent specifically defines the term this way. *See id.*, 8:50-54.

E. Sequence Tag

34. A number of the claims in the '415 patent also recite the term "sequence tag." Ex. 1001, 33:53-36:32. I understand that the Petitioner has offered the broadest reasonable construction of the term "sequence tag" consistent

with the specification as "a DNA sequence of sufficient length that it may be assigned specifically to one of chromosomes 1-22, X or Y." I have used this construction in my analysis and agree with it because the '415 patent specifically defines the term this way. *See id.*, 8:54-56.

F. Massively Parallel Sequencing

35. Claims 5 and 13 recite the term "massively parallel sequencing." Ex. 1001, 35:4-5; 36:1-17. I understand that the Petitioner has offered the broadest reasonable construction of the term "massively parallel sequencing" consistent with the specification as "any technique available as of the effective filing date of the '415 patent for sequencing millions of fragments of nucleic acids." I have used this construction in my analysis and agree with it because the specification of the '415 patent defines the term this way:

- "'Massively parallel sequencing' means techniques for sequencing millions of fragments of nucleic acids, e.g., using attachment of randomly fragmented genomic DNA to a planar, optically transparent surface and solid phase amplification to create a high density sequencing flow cell with millions of clusters, each containing ~1,000 copies of template per sq. cm." Ex. 1001, 9:19-25.
- "These templates are sequenced using four-color DNA sequencing-by-synthesis technology. See, products offered by Illumina, Inc., San

Diego, Calif. In the present work, sequences were obtained, as described below, with an Illumina/Solexa 1G Genome Analyzer." *Id.*, 9:25-29.

G. Mixed Sample

36. A number of the claims in the '415 patent refer to a "mixed sample." Ex. 1001, 33:53-36:32. I understand that the Petitioner has offered the broadest reasonable construction of the term "mixed sample" consistent with the specification as "a sample containing DNA from two different populations, e.g., DNA from a mother and a fetus, or DNA from normal and tumor cells." I have used this construction in my analysis and agree with it because the specification of the '415 patent supports such an interpretation:

- "[T]he present invention comprises, in certain aspects, a method of testing for an abnormal distribution of a specified chromosome portion in a mixed sample of normally and abnormally distributed chromosome portions obtained from a single subject, such as a mixture of fetal and maternal DNA in a maternal plasma sample." Ex. 1001, 4:29-34.
- "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." *Id.*, 4:46-50.

VI. Certain References Teach All of the Claimed Features of the '415 Patent

A. *Lo II* Discloses All of the Features of Claims 1-6 and 8-12 of the '415 Patent

37. In my opinion, as shown in the charts below, *Lo II* discloses each and every feature recited in claims 1-6 and 8-12.

1. Claim 1

38. *Lo II* discloses each and every feature of claim 1.

Claim Language	*Lo II*
1. A method of testing for an abnormal distribution of a specified chromosome portion in a mixed sample of normally and abnormally distributed chromosome portions obtained from a subject, comprising:	*Lo II* discloses methods "for determining whether a nucleic acid sequence imbalance (e.g., chromosome imbalance) exists within a biological sample obtained from a pregnant female." Ex. 1002, [0014]. *Lo II* also discloses that the "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

Claim Language	*Lo II*
	mappable sequenced tags of the specimen." Ex. 1002, [0067]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained in the biological sample, and that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054].
(a) sequencing DNA from the mixed sample to obtain sequences from multiple chromosome portions, wherein said sequences comprise a number of sequence tags of sufficient length of determined sequence to be assigned to a chromosome location with a genome;	*Lo II* discloses that "[a] portion of the nucleic acid molecules contained in the biological sample are sequenced." Ex. 1002, [0015]. *Lo II* also explains that "at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced[,]" and "the nucleic acid molecules are fragments of respective chromosomes." Ex. 1002, [0055]. *Lo II* discloses that the sequencing is done at random. That is, "[t]he origin of a particular fragment is not selected ahead of time." Ex. 1002, [0080]. Because

Claim Language	*Lo II*
	"[t]he sequencing is done at random … a database search may be performed to see where a particular fragment is coming from[,]" indicating that the sequence tag must be of sufficient length to assign the sequence to a location on chromosome a of the genome. Ex. 1002, [0080].
(b) assigning the sequence tags to corresponding chromosome portions including at least the specified chromosome by comparing the determined sequence of the sequence tags to a reference genomic sequence;	*Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II* discloses that "[a]fter the massively parallel sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from." Ex. 1002, [0080]

Claim Language	*Lo II*
(c) determining values for numbers of sequence tags mapping to chromosome portions by using a number of windows of defined length within normally and abnormally distributed chromosome portions to obtain a first value and a second value therefrom; and	*Lo II* discloses, in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes: "In one example, a proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosomes, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allowing the detection of trisomy 21." Ex. 1002, [0069].

Claim Language	*Lo II*
	Lo II discloses, in the same context, that particular "chromosomal regions" are distinct from chromosomes: "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." Ex. 1002, [0060]. *Lo II* discloses using chromosomal regions, or sets of chromosomal regions, to determine if aneuploidy exists: "[t]his determination [of increase or decrease of a clinically-relevant chromosomal region] may be done by using 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. Nucleic acid molecules of the biological sample are

Claim Language	*Lo II*
	sequenced, such that a fraction of the genome is sequenced, and the amount may be determined from results of the sequencing. 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)." Ex. 1002, [0050]. *Lo II* states: "The change detected in the reference quantity may be any deviation (upwards or downwards) in the relation of the clinically-relevant nucleic acid sequence to the other non-clinically-relevant sequences. Thus, the reference state may be any ratio or other quantity (e.g. other than a 1-1 correspondence), and a measured state signifying a change may be any ratio or other quantity that differs from the reference quantity as determined by the one or more cutoff values." Ex. 1002, [0051].

Claim Language	*Lo II*
	Lo II also discloses that "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." Ex. 1002, [0067].
(d) using the values from step (c) to determine a differential, between the first value and the second value, which is determinative of whether or not the abnormal distribution exists.	*Lo II* discloses using the sequencing results to determine first and second amounts of sequences identified as originating from a first and a second chromosome. From those amounts, "[a] parameter from the first amount and the second amount is then compared to one or more cutoff values. Based on the comparison, a classification of whether a fetal chromosomal aneuploidy exists for the first chromosome is determined." Ex. 1002, [0016].

2. Claim 2

39. *Lo II* discloses each and every feature of claim 2.

Claim Language	*Lo II*
2. The method of claim 1 wherein to determine a differential includes the step of comparing a normalized sequence tag density of the specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes: "In one example, a proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosomes, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise." Ex. 1002, [0069]. *Lo II* also discloses deriving from a first amount and a second amount: "[b]ased on the sequencing, a first amount of a first chromosome is determined from sequences identified as originating from the first

Claim Language	*Lo II*
	chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure in found in [0074]: "After this procedure, tags identified as originating from the potentially aneuploid chromosome, i.e. chromosome 21 in this study, are compared quantitatively to all of the sequenced tags or tags originating from one of more chromosomes not involved in the aneuploidy. The relationship between the sequencing output from chromosome 21 and other non-21 chromosomes for a test specimen is compared with cut-off values derived with methods described in the above section to determine if the specimen was obtained from a pregnancy involving a euploid or trisomy 21 fetus." Ex. 1002, [0074].

Claim Language	*Lo II*
	Lo II also states: "[a]lternatively, 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." Ex. 1002, [0075]. Figs. 4A and 4B in *Lo II* show data for all 22 autosomes and the X and Y chromosomes.

3. Claim 3

40. *Lo II* discloses each and every feature of claim 3.

Claim Language	*Lo II*
3. The method of claim 1 wherein the mixed sample comprises a mixture of maternal and fetal DNA and wherein the abnormal distribution results from a fetal aneuploidy.	*Lo II* discloses a "biological sample," which is "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." Ex. 1002, [0033]. "The biological sample may be plasma, urine, serum, or any other suitable sample." Ex. 1002, [0054]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained in the biological sample, and

Claim Language	*Lo II*
	that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054]. *Lo II* discloses an "invention [that] generally relates to the diagnostic testing of fetal chromosomal aneuploidy by determining imbalances between different nucleic acid sequences, and more particularly to the identification of trisomy 21 (Down syndrome) and other chromosomal aneuploidies via testing a maternal sample (e.g. blood)." Ex. 1002, [0003]. *Lo II* also discloses that "[f]etal chromosomal aneuploidy results from the presence of abnormal dose(s) of a chromosome or chromosomal region[,]" which "can be abnormally high, e.g., the presence of an extra chromosome 21 or chromosomal region in trisomy 21." *Id.*, [0004]

4. Claim 4

41. *Lo II* discloses each and every feature of claim 4.

Claim Language	*Lo II*

Claim Language	*Lo II*
4. The method of claim 1 wherein the mixed sample comprises a mixture of normal and genetically altered DNA from a tumor.	*Lo II* discloses that the clinically relevant chromosomal region and the background nucleic acid may come from first and second cell types. According to *Lo II*, "the percentage of fetal sequences in a sample may be determined by any fetal-derived loci and not limited to measuring the clinically-relevant nucleic acid sequences." Ex. 1002, [0052]. *Lo II* further states that "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.* *Lo II* also discloses as "clinically relevant nucleic acid sequences" nucleic acid "sequences which are mutated, deleted, or amplified in a malignant tumor, e.g. sequences in which loss of heterozygosity or gene duplication occur." Ex. 1002, [0037].

5. Claim 5

42. *Lo II* discloses each and every feature of claim 5.

Claim Language	*Lo II*
5. The method of claim 3 wherein the sequencing is massively parallel sequencing.	*Lo II* discloses, as one embodiment, performing the sequencing employed in the aneuploidy detection methods using massively parallel sequencing, which "allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion." Ex. 1002, [0056]. The Illumina Genome Analyzer (or Solexa platform) is identified by *Lo II* as a suitable instrument for performing massively parallel sequencing. *Id.*

6. Claim 6

43. *Lo II* discloses each and every feature of claim 6.

Claim Language	*Lo II*
6. The method of claim 3 wherein the fetal aneuploidy is an aneuploidy of at least one of chromosome 13, 18	*Lo II* discloses that "a parameter (e.g. a fractional representation) of a chromosome potentially involved in a chromosomal aneuploidy, e.g. chromosome 21 or chromosome 18 or chromosome 13, may then be calculated from the results of the bioinformatics

Claim Language	*Lo II*
and 21.	procedure." Ex. 1002, [0063]. Moreover, claim 5 in *Lo II* recites chromosomes 21, 18, and 13 as the chromosomes for which aneuploidy is being tested. Ex. 1002, page 11.

7. Claim 8

44. *Lo II* discloses each and every feature of claim 8.

Claim Language	*Lo II*
8. The method of claim 3 wherein the sequence tags are about 25-100 bp in length.	*Lo II* exemplifies generating sequence tags that are 36 bp in length, (Ex. 1002, [0111]), which is a species within the range of 25-100 bp.

8. Claim 9

45. *Lo II* discloses each and every feature of claim 9.

Claim Language	*Lo II*
9. The method of claim 8 wherein at least about 1	*Lo II* discloses that "[a]s a high number of sequencing reads, in the order of hundred thousands to millions or

Claim Language	*Lo II*
million sequence tags are obtained.	even possibly hundreds of millions or billions, are generated from each sample in each run, the resultant sequenced reads form a representative profile of the mix of nucleic acid species in the original specimen." Ex. 1002, [0057]. In addition, Figs. 6 and 8 in Lo II identify samples having more than one million sequenced tags. Ex. 1002, Figs. 6 and 8.

9. Claim 10

46. *Lo II* discloses each and every feature of claim 10.

Claim Language	*Lo II*
10. The method of claim 8 further comprising the step of calculating a normalized sequence tag density of the specified DNA chromosome portion and a normalized sequence tag density of another DNA	*Lo II* discloses that "a proportion of such sequences [referred to in [0067]] would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosomes, one could obtain a normalized

Claim Language	*Lo II*
chromosome portion in said mixed sample.	frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allowing the detection of trisomy 21. The degree of change in the normalized frequency will be dependent on the fractional concentration of fetal nucleic acids in the analyzed sample." Ex. 1002, [0069]. *Lo II* also discloses that "[o]ne 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)." Ex. 1002, [0014].

10. Claim 11

47. *Lo II* discloses each and every feature of claim 11.

Claim Language	*Lo II*
11. The method of claim 10 wherein the calculating a differential includes the step of comparing a normalized sequence tag density of the specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes: "In one example, a proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosomes, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise." Ex. 1002, [0069]. *Lo II* also discloses deriving a parameter from a first amount and a second amount: "[b]ased on the sequencing, a first amount of a first chromosome is

Claim Language	*Lo II*
	determined from sequences identified as originating from the first chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure in found in [0074]: "After this procedure, tags identified as originating from the potentially aneuploid chromosome, i.e. chromosome 21 in this study, are compared quantitatively to all of the sequenced tags or tags originating from one of more chromosomes not involved in the aneuploidy. The relationship between the sequencing output from chromosome 21 and other non-21 chromosomes for a test specimen is compared with cut-off values derived with methods described in the above section to determine if the specimen was obtained from a pregnancy involving a euploid or trisomy 21 fetus."

Claim Language	*Lo II*
	Ex. 1002, [0074]. *Lo II* also states: "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." Ex. 1002, [0075]. Figs. 4A and 4B in *Lo II* show data for all 22 autosomes and the X and Y chromosomes.

11. Claim 12

48. *Lo II* discloses each and every feature of claim 12.

Claim Language	*Lo II*
12. The method of claim 11 further comprising the step of measuring over- and under-representation of a chromosome by determining a sequence tag density for each	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses deriving a parameter from a first amount and a second amount: "[b]ased on the sequencing, a first amount of a first

Claim Language	*Lo II*
chromosome in the sample, namely chromosomes 1-22, X and also chromosome Y if present.	chromosome is determined from sequences identified as originating from the first chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure in found in [0074]. Ex. 1002, [0074]. *Lo II* states: "The change detected in the reference quantity may be any deviation (upwards or downwards) in the relation of the clinically-relevant nucleic acid sequence to the other non-clinically-relevant sequences. Thus, the reference state may be any ratio or other quantity (e.g. other than a 1-1 correspondence), and a measured state signifying a change may be any ratio or other quantity that differs from the reference quantity as determined by the one

Claim Language	*Lo II*
	or more cutoff values." Ex. 1002, [0051]. Figs. 4A and 4B of *Lo II* show data for all 22 autosomes and the X and Y chromosomes.

B. *Lo II* and *Hillier* and/or *Smith* Teach All of the Features of Claim 7 of the '415 patent

49. In my opinion, *Lo II* and *Hillier* and/or *Smith* teach all of the features recited in claim 7.[1]

50. Claim 7 recites "[t]he method of claim 3 wherein the step of assigning sequence tags to corresponding chromosome portions allows one mismatch."

51. As explained above, *Lo II* discloses each and every feature of claim 3. The disclosure of *Lo II* includes assigning sequence tags to chromosome regions. Ex. 1002, [0014], [0070], [0074], [0080]. *Lo II* is silent as to whether one mismatch is allowed between the sequence tags and the corresponding chromosome portions. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have allowed for one mismatch when assigning sequence tags to corresponding chromosome portions. Doing so is

[1] Charts showing how different combinations of references teach all of the features of the recited claims are attached as Appendices A-O.

merely a known technique to improve similar methods in the same way and yields predictable results.

52. It was well known at the time of the invention of the '415 patent that single nucleotide polymorphisms exist in human DNA sequences obtained from different individuals. It was also known that sequencing methods were not perfect and that errors can exist in sequence tag information. Consequently, a person of ordinary skill in the art would have understood that methods of aligning a sequence tag to a reference sequence should account for these nucleotide differences/errors. For example, *Hillier* discloses the utility of massively parallel short read sequencing for whole genome resequencing and for accurate discovery of genome-wide polymorphisms. Ex. 1006, Abstract. *Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1006, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1006, page 185, Figure 2. In addition, *Smith* teaches that allowing mismatches when mapping sequences to a reference sequence can improve the accuracy of the mapping. Ex. 1009, page 4.

53. Based at least on this knowledge, in my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have permitted one mismatch in sequence tags of sufficient length to assign to a chromosome portion when aligning sequence tags obtained by sequencing DNA from a

biological sample to corresponding chromosome portions of a reference sequence. A person of ordinary skill in the art would have done so to account for the known existence of polymorphisms and sequence errors, thereby increasing the number of usable sequence tags obtained from sequencing the DNA in the sample. Furthermore, a person of ordinary skill would have known that allowing one mismatch still permits one to accurately assign the sequence tag to its corresponding chromosome portion.

C. *Lo II* and *Wang* Teach All of the Features of Claims 13 and 16 of the '415 patent

1. Claim 13

54. In my opinion, *Lo II* and *Wang* teach all of the features recited in claim 13.

a) "A method of determining an abnormally distributed chromosome portion of interest in a mixed sample of normally and abnormally distributed DNA molecules, comprising:"

55. *Lo II* discloses methods "for determining whether a nucleic acid sequence imbalance (e.g., chromosome imbalance) exists within a biological sample obtained from a pregnant female." Ex. 1002, [0014]. *Lo II* goes on to disclose that the "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." Ex. 1002, [0067].

56. *Lo II* discloses a "biological sample," which is "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." Ex. 1002, [0033]. "The biological sample may be plasma, urine, serum, or any other suitable sample." Ex. 1002, [0054]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained in the biological sample, and that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054].

b) "(a) sequencing DNA in said sample by massively parallel sequencing to obtain a number of sequence tags;"

57. *Lo II* discloses that "[a] portion of the nucleic acid molecules contained in the biological sample are sequenced." Ex. 1002, [0015]. *Lo II* also explains that "at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced[,]" and "the nucleic acid molecules are fragments of respective chromosomes." Ex. 1002, [0055]. *Lo II* discloses that the sequencing is done at random. That is, "[t]he origin of a particular fragment is not selected ahead of time." Ex. 1002, [0080]. Because "[t]he sequencing is done at random ... a database search may be performed to see where a particular fragment is coming from[,]" indicating that the sequence tag

must be of sufficient length to assign the sequence to a location on a chromosome of the genome. Ex. 1002, [0080].

58. *Lo II* discloses, as one embodiment, performing the sequencing employed in the aneuploidy detection methods using massively parallel sequencing, which "allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion." Ex. 1002, [0056]. The Illumina Genome Analyzer (or Solexa platform) was identified by *Lo II* as a suitable instrument for performing massively parallel sequencing. *Id.*

c) "(b) mapping said sequence tags to specific chromosome portions, each chromosomal portion being comprised in a sliding window of a predetermined length;"

59. *Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II* discloses that "[a]fter the massively parallel sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from." Ex. 1002, [0080]. *Lo II* does not disclose chromosome portions comprised of a sliding window of a predetermined length.

60. The use of sliding windows in quantitative sequence analyses for the detection of chromosomal aneuploidy was well known in the art at the time of the invention. For example, *Wang* discloses a digital karyotyping method "that provides quantitative analysis of DNA copy number at high resolution." Ex. 1005, Abstract. The method involves first obtaining short sequence tags (21 bp each) from specific locations in the genome. Ex. 1005, page 16156. These tags "generally contain sufficient information to uniquely identify the genomic loci from which they were derived." *Id.* "Second, populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." *Id.* Such a method "can accurately identify regions whose copy number is abnormal." Ex. 1005, page 16161. *Wang* further discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. Depending on the purpose of analysis, e.g., whole chromosome, chromosome arms, amplifications, and deletions, the size of the windows can be different, such as about 4 MB, 200 kb, and 600 kb. Ex. 1005, page 16158, Table 1. Tag densities in a test cell can also be normalized to the tag densities of a reference cell in the same sliding windows. Ex. 1005, page 16159, Figure 2.

61. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have used the known sliding windows sequence data analysis method to normalize sequence tag densities mapped to a reference chromosome with a reasonable expectation of success. Applying this technique to their data analysis, *Wang* detected "[w]hole chromosome changes, gains or losses of chromosomal arms, and interstitial amplification or deletions...." Ex. 1005, page 16161. In my opinion, using *Wang*'s sliding window analysis in the methods of *Lo II* amounts to nothing more than using a known technique to improve similar methods in the same way and yields nothing more than the predictable results.

d) "(c) determining numbers of sequence tags mapped to each sliding window on at least each autosome;"

62. *Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II* discloses that "[a]fter the massively parallel sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from." Ex. 1002, [0080].

63. *Lo II* discloses, in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also

discloses, in the same context, that particular "chromosomal regions" are distinct from chromosomes: "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." Ex. 1002, [0060].

64. *Lo II* discloses using chromosomal regions, or sets of chromosomal regions, to determine if aneuploidy exists: "[t]his determination [of increase or decrease of a clinically-relevant chromosomal region] may be done by using 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. Nucleic acid molecules of the biological sample are sequenced, such that a fraction of the genome is sequenced, and the amount may be determined from results of the sequencing. 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)." Ex. 1002, [0050].

65. *Lo II* also discloses that "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." Ex. 1002, [0067]. *Lo II* also discloses random sequencing a representative fraction of DNA molecules in a sample and then analyzing the chromosomal regions to which they align: "[t]he 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." Ex. 1002, [0108].

66. Among other things, *Wang* discloses that "populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." Ex. 1005, page 16156.

67. *Wang* discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. *Wang* discloses using a sliding windows analysis in methods of digital karyotyping which can detect, among other things, whole chromosome changes. *Wang* discloses using the method to order sequence tags along each chromosome. *Id.* In my opinion, a person having ordinary skill in the art at the time of the invention

would have used sliding windows on tags ordered along each chromosome as taught by *Wang* in the methods of *Lo II* with a reasonable expectation of success as *Wang* discloses using this approach to successfully identify regions of chromosome amplification and deletion. A person of ordinary skill in the art would have done so given this known technique improves the precision of the method (disclosed by *Wang*), which is similar to the methods disclosed in *Lo II*, by allowing normalization to account for differences in local sequence context.

e) "(d) determining a mean of said numbers for each autosome and a second mean for at least all autosomes;"

68. This language in claim 13 requires determining, for each autosome (e.g., human chromosomes 1-22), a mean of the sequence tags in each sliding window for each chromosome, and then calculating a "second mean" that is a mean of the 22 individual means. *Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have used the individual means from each chromosome to calculate a second mean as a method for normalizing the data obtained from all of the sequenced tags mapped to the chromosome portions.

f) "(e) calculating a normalized value from all autosomes, using said second mean; and"

69. *Lo II* discloses normalizing sequence tag density data to account for differences in the relative sizes of chromosomes. Ex. 1002, [0069]. *Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have used the second mean (i.e., the mean of the individual means) to calculate a normalized value for all 22 autosomes because the calculation of normalized values is a standard statistical methodology for adjusting values measured on different scales (in the contest of the claimed methods, sequenced tag densities measured on chromosomes of different sizes) to a notionally common scale. A person of ordinary skill in the art would have known that application of these statistical methods would improve the conclusions drawn from the sequenced tag density data, as demonstrated by the use of averaging normalization of sliding windows data disclosed by *Wang*.

g) "(f) comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest."

70. *Lo II* discloses using the sequencing results to determine first and second amounts of sequences identified as originating from a first and a second

chromosome. From those amounts, "[a] parameter from the first amount and the second amount is then compared to one or more cutoff values. Based on the comparison, a classification of whether a fetal chromosomal aneuploidy exists for the first chromosome is determined." Ex. 1002, [0016]. *Lo II* also states that "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." Ex. 1002, [0075]. Figs. 4A and 4B in *Lo II* show data for all 22 autosomes and the X and Y chromosomes.

71. *Wang* also discloses using normalized sequence tag densities evaluated over moving windows to detect chromosomal aberrations. Ex. 1005, page 16157, and Fig. 1. In addition, *Wang* discloses a comparison of chromosome number analysis for all 22 human autosomes and also the X and Y chromosome. Ex. 1005, page 16158-59, Table 2.

2. Claim 16

72. In my opinion, *Lo II* and *Wang* teach each and every feature recited in claim 16.

73. Claim 16 recites "[t]he method of claim 13 further comprising the step of calculating a normalized value for chromosome X and, if present, Y."

74. As explained above, *Lo II* and *Wang* teach all of the features in claim 13. As just mentioned, *Wang* teaches using normalized values for X and Y

chromosomes. Ex. 1005, page 16158-59, Table 2. *Lo II* discloses normalizing sequence tag densities and mapping sequence tags to chromosomes X and Y. Ex. 1002, [0069], Figs. 4A and 4B. In view of these disclosures in *Lo II* and *Wang,* a person of ordinary skill in the art at the time of the invention of the '415 patent would have calculated normalized values for sequence tags that map to chromosomes X and Y.

D. *Lo II, Shimkets*, and/or *Dohm* Teach All of the Features of Claim 14 of the '415 patent

75. Claim 14 recites "[t]he method of claim 3 further comprising the step of calculating a relationship between numbers of sequence tags and GC content associated with sequence tags in a given window and correcting for a higher or lower number of reads resulting from a change in GC content."

76. As explained above, *Lo II* teaches each and every feature of claim 3. *Shimkets* is directed to sequence-based karyotyping. *Shimkets* discloses that "inherent in the sequencing process itself may be a slight bias in favor of sequences with certain compositional characteristics (such as higher or lower GC content, the percentage of nucleotides in a given stretch that are G or C)." Ex. 1004, ¶ [0075]. *Shimkets* teaches that "[t]his bias could be ascertained by calibration experiments and then factored in to subsequent computationally derived reference distributions." *Id.*

77. *Dohm* observed "a strong correlation between GC richness and read coverage, with the read density being increased in regions of elevated GC content" for the Solexa sequencing platform. Ex. 1007, page e104. "Thus, Solexa-based de novo sequencing as well as re-sequencing activities need to calibrate their sequencing output for achieving accordingly high read coverage of AT-rich regions." Ex. 1007, page e105.

78. From the teaching of *Shimkets*, a person of ordinary skill in the art knew that GC content can bias sequencing results and accordingly that bias could be accounted for in evaluating sequence data. *Dohm* confirms that the GC bias is present in sequence read coverage in data obtained from the Illumina/Solexa massively parallel sequencing (MPS) technology. Knowing of the potential for GC bias to have an impact on sequence tag densities, in my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have applied *Shimkets*' and/or *Dohm*'s disclosure of accounting for the bias when analyzing karyotyping data to the methods disclosed in *Lo II* with a reasonable expectation of success at the time of the invention.

E. *Lo II* and *Quake* Teach All of the Features of Claim 15 of the '415 patent

79. Claim 15 recites "[t]he method of claim 3 further comprising the step of calculating a t statistic for each chromosome relative to other chromosomes in

the mixed sample, whereby each t statistic indicates a value of a chromosome relative to other chromosomes in a sample, said value being indicative of disomy."

80. As explained above, *Lo II* teaches all of the features of claim 3. The use of t statistics in data analysis is conventional in the art. For example, *Quake* discloses that a t-statistic is a statistical method known in the art. Ex. 1008, 5:64-67 ("A commonly used measure of statistical significance when a highly significant result is desired is $p<0.01$, i.e., a 99% confidence interval based on a chi-square or t-test."). In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have applied conventional statistical analyses, such as a t-test statistic, to the methods disclosed in *Lo II* with a reasonable expectation of success. A person of ordinary skill in the art would have been motivated to use the confidence intervals derived from t statistics when evaluating sequence tag density data to determine the disomy of chromosomes in a mixed sample.

F. *Lo II, Wang,* and *Hillier* and/or *Smith* Teach All of the Features of Claim 17 of the '415 patent

81. Claim 17 recites "[t]he method of claim 13 wherein said mapping includes mapping sequences with one mismatch."

82. As explained above, *Lo II* and *Wang* teach all of the features of claim 13. These references are silent as to whether one mismatch is allowed between the sequence tags and the corresponding chromosome portions. In my opinion, a

person of ordinary skill in the art at the time of the invention of the '415 patent would have allowed for one mismatch when assigning sequence tags to corresponding chromosome portions. Doing so is merely a known technique to improve similar methods in the same way and yields predictable results.

83. As explained above, it was well known at the time of the invention that single nucleotide polymorphisms exist in human DNA sequences obtained from different individuals. It was also known that sequencing methods were not perfect and that errors can exist in sequence tag information. Consequently, methods of aligning a sequence tag to a reference sequence should account for these nucleotide differences/errors. *Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1006, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1006, page 185, Figure 2. In addition, *Smith* teaches that allowing mismatches when mapping sequences to a reference sequence can improve the accuracy of the mapping. Ex. 1009, page 4.

84. Based at least on this knowledge, in my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have permitted one mismatch in sequence tags of sufficient length to assign to a chromosome portion when aligning sequence tags obtained by sequencing DNA from a biological sample to corresponding chromosome portions of a reference sequence.

A person of ordinary skill in the art would have done so to account for the known existence of polymorphisms and sequence errors, thereby increasing the number of usable sequence tags obtained from a sequencing the DNA in the sample. *Id.* Furthermore, a person of ordinary skill would have known that allowing one mismatch still permits one to assign the sequence tag to its corresponding chromosome portion. *Id.* Therefore, in my opinion, *Lo II*, *Wang*, and *Hillier* and/or *Smith* teach all of the features in claim 17.

G. *Lo II* and *Wang* Teach All of the Features of Claims 1-6 and 8-12 of the '415 patent

85. As explained above, *Lo II* discloses all of the features in claims 1-6 and 8-12. Claims 1-6 and 8-12 are directed to methods that include "using a number of windows of defined length within normally and abnormally distributed chromosome portions." Ex. 1001, claim 1. In my opinion, both *Lo II* and *Wang* disclose this feature, although *Wang* discloses this feature of the claims in more detail. I am of the opinion that a person of ordinary skill in the art at the time of the alleged invention of the '415 patent would also have modified *Lo II*'s methods to include this feature in view of the more detailed disclosure in *Wang*. Doing so amounts to nothing more than using a known technique to improve *Lo II*'s methods in the same way as the use of windows improves *Wang*'s methods, and yields nothing more than predictable results.

86. Among other things, *Wang* discloses that "populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." Ex. 1005, page 16156. And as mentioned, *Wang* discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. *Wang* discloses using a sliding windows analysis in methods of digital karyotyping which can detect, among other things, whole chromosome changes.

87. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have used sliding windows on tags ordered along each chromosome as taught by *Wang* in the methods of *Lo II* with a reasonable expectation of success as *Wang* discloses using this approach to successfully identify regions of chromosome amplification and deletion. A person of ordinary skill in the art would have done so given this known technique improves methods (disclosed by *Wang*) that are similar to the methods disclosed in *Lo II*.

H. *Lo II, Wang,* and *Hillier* and/or *Smith* Teach All of the Features of Claim 7 of the '415 patent

88. As explained above, *Lo II* and *Hillier* and/or *Smith* teach all of the features in claim 7. As also explained above, *Wang* discloses using sliding

windows in methods for detecting chromosome aberrations. Given this disclosure by *Wang*, I am also of the opinion that a person of ordinary skill in the art at the time of the invention of the '415 patent would have combined the teachings of *Lo II*, *Wang*, and *Hillier* and/or *Smith* to arrive at the invention of claim 7.

I. *Lo II, Wang, Shimkets,* and/or *Dohm* Teach All of the Features of Claim 14 of the '415 patent

89. As explained above, *Lo II*, *Shimkets*, and/or *Dohm* teach all of the features in claim 14. As also explained above, *Wang* discloses using sliding windows in methods for detecting chromosome aberrations. Given this disclosure by *Wang*, I am also of the opinion that a person of ordinary skill in the art at the time of the invention of the '415 patent would have combined the teachings of *Lo II*, *Wang*, *Shimkets*, and/or *Dohm* to arrive at the invention of claim 14.

J. *Lo II, Wang,* and *Quake* Teach All of the Features of Claim 15 of the '415 patent

90. As explained above, *Lo II* and *Quake* teach all of the features in claim 15. As also explained above, *Wang* discloses using sliding windows in methods for detecting chromosome aberrations. Given this disclosure by *Wang*, I am also of the opinion that a person of ordinary skill in the art at the time of the invention of the '415 patent would have combined the teachings of *Lo II*, *Wang*, and *Quake* to arrive at the invention of claim 15.

K. *Lo I* and *Shimkets* Teach All of the Features of Claims 1-6 and 8-12 of the '415 patent

91. In my opinion, *Lo I* and *Shimkets* teach all of the features recited in claims 1-6 and 8-12.

92. *Lo I* provides, among other things, methods "for determining whether a nucleic acid sequence imbalance (e.g., allelic imbalance) exists within a biological sample." Ex. 1003, [0010]. A "biological sample" is "any sample that is taken from a subject (*e.g.*, a human, such as a pregnant woman) and contains one or more nucleic acid sof [*sic*] of interest." Ex. 1003, [0030]. The biological sample may be maternal plasma, which contains fetal nucleic acid sequences and maternal nucleic acid sequences. Ex. 1003, [0044].

93. Paragraph [0192] of *Lo I* discloses the following MPS method for detecting fetal chromosomal aneuploidies:

> "Here we shall describe another example whereby a variant of digital PCR can be used for the detection of fetal chromosomal aneuploidies, using the example of trisomy 21, in maternal plasma. The variant of digital PCR is the performance of massively parallel genomic sequencing using emulsion PCR in a sequencing machine such as the Roche GS20 system (http://www.454.com/about-454/partners.asp) the Applied Biosystems 'supported oligo ligation detection' (SOLiD) and the Illumina Solexa sequencing technology.

The general principle of this strategy is that if one is to do random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother. **A proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosome, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allow [*sic*] the detection of trisomy 21. The degree of change in the normalized frequency will be dependent on the fractional concentration of fetal nucleic acids in the analyzed sample.** It should be obvious to those of skill in the art that a proportion of the sequencing results will come from repetitive sequences which might be difficult to be attributed to individual chromosomes but appropriate

statistical analysis can be performed to take this fact into consideration."

Ex. 1003, [0192]; bold text appears in *Lo II*, Ex. 1002 [0069].

1. Claim 1

94. *Lo I* and *Shimkets* teach each and every feature recited in claim 1.

a) "A method of testing for an abnormal distribution of a specified chromosome portion in a mixed sample of normally and abnormally distributed chromosome portions obtained from a subject, comprising:"

95. *Lo I* discloses a method for detecting fetal chromosomal aneuploidies, using the example of trisomy 21, by performing random sequencing of DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192]. The DNA fragments, or genomic sequences, would have originally come from either the fetus or the mother. Ex. 1003, [0192]. In other words, *Lo I* discloses a method for testing for an abnormal distribution of a specified chromosome portion (e.g., chromosome 21) in a mixed sample containing normally and abnormally distributed chromosome portions.

b) "(a) sequencing DNA from the mixed sample to obtain sequences from multiple chromosome portions, wherein said sequences comprise a number of sequence tags of sufficient length of determined sequence to be assigned to a chromosome location within a genome;"

96. *Lo I* discloses that one may "do random sequencing of DNA fragments that are present in the plasma of a pregnant woman," and that in doing

so "one would obtain genomic sequences which would originally have come from either the fetus or the mother." Ex. 1003, [0192]. According to *Lo I*, "[a] proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes." Ex. 1003, [0192]. *Lo I* does not expressly state that the sequences (corresponding to the "sequence tags" of claim 1) are "of sufficient length of determined sequence to be assigned to a chromosome location within a genome," but that must necessarily be the case because that is the only way to determine from which chromosomes the random sequences are derived.

97. *Shimkets* discloses a sequence-based karyotyping method that "may be used to determine chromosomal abnormalities including balanced and unbalanced chromosomal rearrangements, polyploidy, aneuploidy, deletions, duplications, copy number polymorphisms and the like." Ex. 1004, [0063]. The method comprises "generating a pool of fragments of genomic DNA by a random fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution expected by chance." Ex. 1004, [0007]; Figure 9. The at least 20 contiguous bases obtained "will typically allow the mapping of the

fragment to a unique location in a genomic scaffold." Ex. 1004, [0071]. Thus, *Shimkets* expressly discloses generating sequence tags of sufficient length to uniquely assign them to a chromosome location in a genome.

98. Unlike *Lo I*, which teaches using a mixed sample that includes cell free maternal and fetal DNA, *Shimkets* teaches performing digital karyotyping on separate samples, for example, a reference "normal" cell sample and a test cell sample from an individual suspected of having cancer. The results obtained by sequencing these samples are normalized by the application of sequence analyses and statistical methods that are conventional in the art. *See, e.g.*, Ex. 1004, [0007], [0012], [0073], [0267]. In my opinion, these well-known methods of sequence/statistical analyses are equally applicable to the sequence data obtained from *Shimkets'* individually sequenced samples as they are to *Lo I*'s sequenced mixed sample. In other words, there is nothing unique in *Shimkets* teaching of normalizing data, and the disclosure in *Lo I* that the sequencing data from the mixed samples may be normalized would suggest to a person of ordinary skill in the art at the time of the invention to utilize the data normalization methods disclosed by *Shimkets*.

c) **"(b) assigning the sequence tags to corresponding chromosome portions including at least the specified chromosome by comparing the determined sequence of the sequence tags to a reference genomic sequence;"**

99. Lo I discloses a method in which "[t]he general principle … is that if one is to do random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother. A proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes." Ex. 1003, [0192]. Lo I discloses obtaining the sequences and then using them to determine a normalized frequency by taking into account the relative sizes of the chromosomes from which the sequences were derived. Id. Lo I does not expressly disclose assigning the obtained genomic sequences to corresponding chromosome portions, including the specified chromosome (e.g., chromosome 21) by comparing the obtained genomic sequences (i.e., sequence tags) to a reference genomic sequence. But one of ordinary skill in the art reading Lo I would know that once "genomic sequences" had been obtained from random sequencing of DNA fragments from a maternal plasma sample, the only way to assign those sequences to chromosome 21, or to other chromosomes, would be by comparing the "genomic sequences" to a reference genomic sequence.

100. In addition, *Shimkets* discloses "generating a pool of fragments of genomic DNA by a random fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution expected by chance." Ex. 1004, ¶ [0007]; Figure 9. A person of ordinary skill in the art at the time of the invention of the '415 patent would have known that comparing the obtained "genomic sequences" disclosed in *Lo I* to a reference genome (*Shimkets*' "genomic scaffold") is the same as assigning sequence tags to their corresponding chromosome portions as recited in the claim.

d) **"(c) determining values for numbers of sequence tags mapping to chromosome portions by using a number of windows of defined length within normally and abnormally distributed chromosome portions to obtain a first value and a second value therefrom; and"**

101. *Lo I* discloses normalizing the data obtained from the mapped sequences to account for differences in the respective sizes of different chromosomes. Ex. 1003, [0192] ("By taking into account of the relative size of chromosome 21 compared with the other chromosome, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the

normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allow [*sic*] the detection of trisomy 21.").

102. *Shimkets* discloses normalizing the data obtained from mapped sequences. The "[r]atios, on a per chromosomal basis, of the number of uniquely mapping fragments in the experimental sample to the number in the normal sample (corrected by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples) can be used to estimate rates of aneuploidy." Ex. 1004, ¶ [0267].

103. *Shimkets* also discloses normalizing data by obtaining the distribution of the fragments using a number of windows of defined length within a test chromosome (either normal or abnormal) and a normal chromosome. "The number of a plurality of sequences mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, ¶ [0007].

104. *Shimkets* further explains the concept of "windows" in relation to "the test cell distribution (i.e., chromosomal map density)," which "is defined as the number of mapped sequences (i.e., fragments) by the number of possible map locations present in a given chromosome. The number of possible map locations is defined by the size of the observation window and the length of the chromosome." Ex. 1004, ¶¶ [0012], [0073].

105. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have normalized the sequence tag data as disclosed by *Lo I* by applying the windows-based normalization approach disclosed by *Shimkets* to determine values for numbers of sequence tags mapping to normally and abnormally distributed chromosome portions to obtain first and second values therefrom. *Lo I* and *Shimkets* are both directed to using random sequencing of nucleic acids to detect chromosomal abnormalities. A person of ordinary skill in the art would recognize that benefits of the windows-based data normalization disclosed by *Shimkets* would be desirable in the methods disclosed by *Lo I*. Moreover, a person of ordinary skill in the art at the time of the invention could implement the windows-based normalization methods disclosed by *Shimkets* in the methods disclosed by *Lo I* with a reasonable expectation of success given that the methods in both references involve mapping sequences to reference

chromosomes (or regions of chromosomes) having either normal or abnormal distributions.

e) **"(d) using the values from step (c) to determine a differential, between the first value and the second value, which is determinative of whether or not the abnormal distribution exists."**

106. As just discussed, *Shimkets* discloses that "[t]he number of a plurality of sequences mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, ¶ [0007]. Thus, *Shimkets* discloses step (d) of '415 patent claim 1.

107. In summary, the combination of *Lo I* and *Shimkets* teaches all of the steps recited in claim 1 of the '415 patent. In my opinion, a person of ordinary skill in the art at the time of the invention would have been motivated to combine these references with a reasonable expectation of success in arriving at the method of testing for an abnormal distribution of a specified chromosome portion recited in claim 1.

2. Claim 2

108. *Lo I* and *Shimkets* teach all of the features recited in claim 2.

109. Claim 2 recites "[t]he method of claim 1 wherein to determine a differential includes the step of comparing a normalized sequence tag density of the specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density."

110. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 1. A person of ordinary skill in the art would apply the data normalization approaches disclosed in *Shimkets* to the sequence data obtained from *Lo I*'s mixed sample, as discussed above. *Shimkets* discloses that the ratio, on a per chromosomal basis, of the number of mapped sequences in an experimental sample to the number in the normal sample can be normalized "by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples." Ex. 1004, [0267]. For instance:

> "[c]ounts of the resulting number of unique hits to each chromosome were tabulated for both the test DiFi sample and the reference GM12911 sample. For each chromosome, the ratio of the number of unique hits in the DiFi sample to the corresponding number of hits to the GM12911 sample was computed, providing a raw ratio of measured chromosomal content on a per chromosome basis. The raw

ratios were further normalized to account for any difference in the amount of actual sequencing performed for the two samples; specifically, the ratio of the total number of unique hits to the autosomal chromosomes in the DiFi and GM12911 samples was used as a multiplicative normalization factor to convert the raw chromosomal content ratios into normalized ratios."

Ex. 1004, [0248].

111. *Shimkets* teaches the data analysis feature recited in claim 2. In view of this disclosure, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 2 based on the combination of *Lo I* and *Shimkets*.

3. Claim 3

112. *Lo I* and *Shimkets* teach each and every feature recited in claim 3.

113. Claim 3 recites "[t]he method of claim 1 wherein the mixed sample comprises a mixture of maternal and fetal DNA and wherein the abnormal distribution results from a fetal aneuploidy."

114. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 1. *Lo I* discloses a method for detecting fetal chromosomal aneuploidies, using the example of trisomy 21, by performing random sequencing of cell free DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192].

The DNA fragments, or genomic sequences, would have originally come from either the fetus or the mother. Ex. 1003, [0192]. Given this disclosure in *Lo I*, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 3 based on the combined disclosures of *Lo I* and *Shimkets*.

4. Claim 4

115. *Lo I* and *Shimkets* teach each and every feature recited in claim 4.

116. Claim 4 recites "[t]he method of claim 1 wherein the mixed sample comprises a mixture of normal and genetically altered DNA from a tumor."

117. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 1. *Shimkets* discloses that "Sequence-Based Karyotyping or high resolution molecular karyotyping according to the invention can be used to identify remaining oncogenes and tumor suppressor genes...." Ex. 1004, [0092]. Shimkets discloses this embodiment as a comparison of "the genomes from a normal subject and a diseased subject." *Id.* *Shimkets* does not disclose using a mixed cell free sample of DNA.

118. *Lo I* describes methods using a mixed sample comprising a mixture of normal and genetically altered DNA. Ex. 1003, [0192]. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have modified the methods of *Shimkets* in view of the disclosure in *Lo I* by

substituting *Lo I*'s mixed sample for *Shimkets*' individual samples, with a reasonable expectation of success, at the time of the invention.

5. Claim 5

119. *Lo I* and *Shimkets* teach each and every feature recited in claim 5.

120. Claim 5 recites "[t]he method of claim 3 wherein the sequencing is massively parallel sequencing."

121. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 3. Both *Lo I* and *Shimkets* disclose methods involving random sequencing using massively parallel genomic sequencing. Ex. 1003, [0192]; Ex. 1004, [0258]. Given this disclosure in *Lo I* and *Shimkets*, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 5 based on the combined disclosures of *Lo I* and *Shimkets*.

6. Claim 6

122. *Lo I* and *Shimkets* teach each and every feature recited in claim 6.

123. Claim 6 recites "[t]he method of claim 3 wherein the fetal aneuploidy is an aneuploidy of at least one of chromosome 13, 18 and 21."

124. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 3. *Lo I* discloses a method for detecting fetal chromosomal aneuploidies, using the example of trisomy 21, by performing random sequencing of DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192]. Given

this disclosure in *Lo I*, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 6 based on the combined disclosures of *Lo I* and *Shimkets*.

7. Claim 8

125. *Lo I* and *Shimkets* teach each and every feature recited in claim 8.

126. Claim 8 recites "[t]he method of claim 3 wherein the sequence tags are about 25-100 bp in length."

127. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 3. *Shimkets* discloses that "[w]hile the sequencing of 20 bp from each fragment is sufficient, sequencing of more bases is also useful. For example, the sequencing of at least 25 bp, at least 30 bp, at least 35 bp, at least 40 bp, at least 45 bp, at least 50 bp, at least 55 bp, at least 60 bp, at least 65 bp, at least 70 bp, at least 75 bp, at least 80 bp, at least 95 bp, at least 100 bp have been performed by the methods of the invention and found to be useful but not essential." Ex. 1004, [0070]. Given this disclosure in *Shimkets*, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 8 based on the combined disclosures of *Lo I* and *Shimkets*.

8. Claim 9

128. *Lo I* and *Shimkets* teach each and every feature recited in claim 9.

129. Claim 9 recites "[t]he method of claim 8 wherein at least about 1 million sequence tags are obtained."

130. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 8. *Shimkets* discloses that "[a]t least 1000, 10,000, 100,000, 1,000,000 or more sequenced are mapped." Ex. 1004, ¶ [0011]. Given this disclosure in *Shimkets*, a person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 9 based on the combined disclosures of *Lo I* and *Shimkets*.

9. Claim 10

131. *Lo I* and *Shimkets* teach each and every feature recited in claim 10.

132. Claim 10 recites "[t]he method of claim 8 further comprising the step of calculating a normalized sequence tag density of the specified DNA chromosome portion and a normalized sequence tag density of another DNA chromosome portion in said mixed sample."

133. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 8. Also as discussed above in the context of claim 2, *Shimkets* discloses that the ratio, on a per chromosomal basis, of the number of mapped sequences in an experimental sample to the number in the normal sample can be normalized "by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for

differences in the amount of sequencing in the two samples." Ex. 1004, ¶ [0267]. *Shimkets* discloses calculating normalized ratios for the autosomal chromosomes from normal (reference GM12911) and abnormal (DiFi) cells. Ex. 1004, ¶ [0248]. *Lo I* teaches using a mixed sample. Ex. 1003, [0192].

10. Claim 11

134. *Lo I* and *Shimkets* teach each and every feature recited in claim 11.

135. Claim 11 recites "[t]he method of claim 10 wherein the calculating a differential includes the step of comparing a normalized sequence tag density of the specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density."

136. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 10. *Lo I* teaches using a mixed sample. Ex. 1003, [0192]. Also as discussed above in the context of claim 2, *Shimkets* discloses that "[t]he raw ratios were further normalized to account for any difference in the amount of actual sequencing performed for the two samples; specifically, the ratio of the total number of unique hits to the autosomal chromosomes in the DiFi and GM12911 samples was used as a multiplicative normalization factor to convert the raw chromosomal content ratios into normalized ratios." Ex. 1004, [0248]. In view of this disclosure, a person of ordinary skill in the art at the time of the invention of

the '415 patent would have arrived at the invention of claim 11 based on the combination of *Lo I* and *Shimkets*.

11. Claim 12

137. *Lo I* and *Shimkets* teach all of the features recited in claim 12.

138. Claim 12 recites "[t]he method of claim 11 further comprising the step of measuring over- and under-representation of a chromosome by determining a sequence tag density for each chromosome in the sample, namely chromosomes 1-22, X and also chromosome Y if present."

139. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 11. *Shimkets* teaches evaluating aneuploidy across the entire genome, including the X and Y chromosomes. For example, *Shimkets* states: "In the extreme, one could make a contingency table of the entire genome, with one column per chromosome to identify chromosomes that are over or underrepresented in content at the entire chromosomal level. Ratios, on a per chromosomal basis, of the number of uniquely mapping fragments in the experimental sample to the number in the normal sample (corrected by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples), can be used to estimate rates of aneuploidy." Ex. 1004, [0267]. In view of this teaching, a person

of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 12 based on the combination of *Lo I* and *Shimkets*.

L. *Lo I, Shimkets,* and *Hillier* and/or *Smith* Teach Each and Every Feature of Claim 7 of the '415 patent

140. Claim 7 recites "[t]he method of claim 3 wherein the step of assigning sequence tags to corresponding chromosome portions allows one mismatch."

141. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 3. These references are silent as to whether one mismatch is allowed between the sequence tags and the corresponding chromosome portions. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have allowed for one mismatch when assigning sequence tags to corresponding chromosome portions.

142. For example, *Hillier* discloses the utility of massively parallel short read sequencing for whole genome resequencing and for accurate discovery of genome-wide polymorphisms. Ex. 1006, Abstract. *Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1006, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1006, page 185, Figure 2. In addition, *Smith* teaches that allowing mismatches when mapping

sequences to a reference sequence can improve the accuracy of the mapping. Ex. 1009, page 4.

143. Based at least on this knowledge, a person of ordinary skill in the art at the time of the invention of the '415 patent would have permitted one mismatch in sequence tags of sufficient length to assign to a chromosome portion when aligning sequence tags obtained by sequencing DNA from a biological sample to corresponding chromosome portions of a reference sequence.

M. *Lo I, Shimkets,* and *Wang* Teach Each and Every Feature of Claims 13 and 16 of the '415 patent

1. Claim 13

144. *Lo I*, *Shimkets*, and *Wang* teach each and every feature recited in claim 13.

a) "A method of determining an abnormally distributed chromosome portion of interest in a mixed sample of normally and abnormally distributed DNA molecules, comprising:"

145. As discussed in the context of claim 1, *Lo I* discloses a method for testing for an abnormal distribution of a specified chromosome portion (e.g., chromosome 21) in a mixed sample containing cell free normally and abnormally distributed chromosome portions. Ex. 1003, [0192].

b) "(a) sequencing DNA in said sample by massively parallel sequencing to obtain a number of sequence tags;"

146. *Lo I* discloses methods involving random sequencing using massively parallel genomic sequencing to obtain a number of "genomic sequences" (i.e., sequence tags). Ex. 1003, [0192]. *Shimkets* also discloses using a massively parallel sequencing platform, a pyrophosphate sequencer from 454 Life Sciences (New Haven, Conn.), which is capable of sequencing 70,000 beads simultaneously. Ex. 1004, [0580].

c) "(b) mapping said sequence tags to specific chromosome portions, each chromosomal portion being comprised in a sliding window of a predetermined length;"

147. *Lo I* discloses a method in which "[t]he general principle ... is that if one is to do random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother. A proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes." Ex. 1003, [0192]. *Lo I* discloses obtaining the sequences and then using them to determine a normalized frequency by taking into account the relative sizes of the chromosomes from which the sequences were derived. *Id.* *Lo I* does not expressly disclose assigning the obtained genomic sequences to corresponding chromosome portions,

including the specified chromosome (e.g., chromosome 21) by comparing the obtained genomic sequences (i.e., sequence tags) to a reference genomic sequence. But one of ordinary skill in the art reading *Lo I* would know that once "genomic sequences" had been obtained from random sequencing of DNA fragments from a maternal plasma sample, the only way to assign those sequences to chromosome 21, or to other chromosomes, would be by comparing the "genomic sequences" to a reference genomic sequence.

148. In addition, *Shimkets* discloses "generating a pool of fragments of genomic DNA by a random fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution expected by chance." Ex. 1004, [0007]; Figure 9. A person of ordinary skill in the art at the time of the invention would have known that comparing the obtained "genomic sequences" disclosed in *Lo I* to a reference genome (*Shimkets*' "genomic scaffold") is the same as assigning sequence tags to their corresponding chromosome portions as recited in the claim.

149. The use of sliding windows in quantitative sequence analyses for the detection of chromosomal aneuploidy was well known in the art at the time of the invention. For example, *Wang* discloses a digital karyotyping method "that

provides quantitative analysis of DNA copy number at high resolution." Ex. 1005, Abstract. The method involves first obtaining short sequence tags (21 bp each) from specific locations in the genome. Ex. 1005, page 16156. "These tags generally contain sufficient information to uniquely identify the genomic loci from which they were derived. Second, populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." *Id.* Such a method "can accurately identify regions whose copy number is abnormal." Ex. 1005, page 16161. *Wang* further discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. Depending on the purpose of analysis, e.g., the whole chromosome, chromosome arms, amplifications, and deletions, the size of the windows can be different, such as about 4 MB, 200 kb, and 600 kb. Ex. 1005, page 16158, Table 1. Tag densities in a test cell can also be normalized to the tag densities of a reference cell in the same sliding windows. Ex. 1005, page 16159, Figure 2.

150. In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have used the known sliding windows sequence data analysis method to normalize sequence tag densities mapped to a reference

chromosome with a reasonable expectation of success. Applying this technique to their data analysis, *Wang* detected "[w]hole chromosome changes, gains or losses of chromosomal arms, and interstitial amplification or deletions...." Ex. 1005, page 16161. Using *Wang*'s sliding window analysis in the methods of *Lo I* amounts to nothing more than using a known technique to improve similar methods in the same way and yields nothing more than the predictable results.

d) "(c) determining numbers of sequence tags mapped to each sliding window on at least each autosome;"

151. *Shimkets* discloses that "[t]he number of a plurality of sequences mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, [0007]. In discussing mapping sequences to chromosomes in the genome, *Shimkets* discloses that "[t]he test cell distribution (i.e., chromosomal map density) is defined as the number of mapped sequences (i.e., fragments) by the number of possible map locations present in a given chromosome. The number of possible map locations is defined by the size of the observation window and the length of the chromosome. Ex. 1004, [0012]. *Lo I* and *Shimkets* do not disclose sliding windows.

152. As explained above, *Wang* discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, 16160. *Wang* discloses using a sliding windows analysis in methods of digital karyotyping which can detect, among other things, whole chromosome changes. A person of ordinary skill in the art at the time of the invention of the '415 patent would have used sliding windows on tags ordered along each chromosome as taught by *Wang* in the methods of *Lo I* and *Shimkets* with a reasonable expectation of success as *Wang* discloses using this approach to successfully identify regions of chromosome amplification and deletion. A person of ordinary skill in the art would have done so given this known technique improves methods (disclosed by *Wang*) that are similar to the methods disclosed in *Lo I* and *Shimkets*.

e) "(d) determining a mean of said numbers for each autosome and a second mean for at least all autosomes;"

153. This language in claim 13 requires determining, for each autosome (e.g., human chromosomes 1-22), a mean of the sequence tags in each sliding window for each chromosome, and then calculating a "second mean" that is a mean of the 22 individual means. *Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157. A person of ordinary skill in the art at the time of the invention of the '415 patent would have used the individual

means from each chromosome to calculate a second mean as a method for normalizing the data obtained from all of the sequenced tags mapped to the chromosome portions.

f) "(e) calculating a normalized value from all autosomes, using said second mean; and"

154. *Lo I* discloses normalizing sequence data taking into account the relative sizes of chromosomes. Ex. 1003, [0192]. *Shimkets* discloses normalizing the number of sequences mapped to different chromosomal regions. Ex. 1004, [0248], [0267].

155. *Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157. A person of ordinary skill in the art at the time of the invention of the '415 patent would have used the second mean (i.e., the mean of the individual means) to calculate a normalized value for all 22 autosomes because the calculation of normalized values is a standard statistical methodology for adjusting values measured on different scales (in the context of the claimed methods, sequenced tag densities measured on chromosomes of different sizes) to a notionally common scale. A person of ordinary skill in the art would have known that application of these statistical methods would improve the conclusions

drawn from the sequenced tag density data, as demonstrated by the use of averaging normalization of sliding windows data disclosed by *Wang*.

g) "(f) comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest."

156. *Lo I* discloses comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest. ("If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allow the detection of trisomy 21."). Ex. 1003, [0192]. Although *Shimkets* does not disclose using a mixed sample, the entirety of *Shimkets'* sequence-based karyotyping method entails making a comparison between normally and abnormally distributed chromosomes in separate samples. In addition, *Wang* also discloses using normalized sequence tag densities evaluated over sliding windows to detect chromosomal aberrations. Ex. 1005, page 16157, and Fig. 1. In addition, *Wang* discloses a comparison of chromosome number analysis for all 22 human autosomes and also the X and Y chromosome. Ex. 1005, page 16158-59, Table 2.

2. Claim 16

157. *Lo I*, *Shimkets*, and *Wang* teach each and every feature recited in claim 16.

158. Claim 16 recites "[t]he method of claim 13 further comprising the step of calculating a normalized value for chromosome X and, if present, Y."

159. As explained above, *Lo I*, *Shimkets*, and *Wang* teach all of the features of claim 13. As just mentioned, *Wang* teaches using normalized values for X and Y chromosomes. Ex. 1005, page 16158-59, Table 2. *Lo I* and *Shimkets* disclose normalizing sequence tag densities and mapping sequence tags to chromosomes X and Y. Ex. 1003, [0192]; Ex. 1004, [0267], [0248]. In view of these disclosures, a person of ordinary skill in the art at the time of the invention of the '415 patent would have calculated normalized values for sequence tags that map to chromosomes X and Y.

N. *Lo I, Shimkets,* and/or *Dohm* Teach Each and Every Feature of Claim 14 of the '415 patent

160. Claim 14 recites "[t]he method of claim 3 further comprising the step of calculating a relationship between numbers of sequence tags and GC content associated with sequence tags in a given window and correcting for a higher or lower number of reads resulting from a change in GC content."

161. As discussed above, *Lo I* and *Shimkets* teach all of the features of claim 3. *Shimkets* is directed to sequence-based karyotyping. *Shimkets* discloses that "inherent in the sequencing process itself may be a slight bias in favor of sequences with certain compositional characteristics (such as higher or lower GC content, the percentage of nucleotides in a given stretch that are G or C)." Ex.

1004, ¶ [0075]. *Shimkets* teaches that "[t]his bias could be ascertained by calibration experiments and then factored in to subsequent computationally derived reference distributions." Ex. 1004, ¶ [0075].

162. *Dohm* observed "a strong correlation between GC richness and read coverage, with the read density being increased in regions of elevated GC content" for the Solexa sequencing platform. Ex. 1007, page e104. "Thus, Solexa-based de novo sequencing as well as re-sequencing activities need to calibrate their sequencing output for achieving accordingly high read coverage of AT-rich regions." Ex. 1007, page e105.

163. From the teaching of *Shimkets*, a person of ordinary skill in the art knew that GC content can bias sequencing results and accordingly the bias could be accounted for in evaluating sequence data. *Dohm* confirms that the GC bias is present in sequence read coverage in data obtained from the Illumina/Solexa MPS technology. Knowing of the potential for GC bias to have an impact on sequence tag densities, a person of ordinary skill in the art at the time of the invention would have applied *Shimkets*' and/or *Dohm's* disclosure of accounting for the bias when analyzing karyotyping data to the methods disclosed in *Lo I* with a reasonable expectation of success.

O. *Lo I, Shimkets,* and *Quake* Teach Each and Every Feature of Claim 15 of the '415 patent

164. Claim 15 recites "[t]he method of claim 3 further comprising the step of calculating a t statistic for each chromosome relative to other chromosomes in the mixed sample, whereby each t statistic indicates a value of a chromosome relative to other chromosomes in a sample, said value being indicative of disomy."

165. As explained above, *Lo I* and *Shimkets* teach all of the features of claim 3. The use of t statistics in data analysis is conventional in the art. For example, *Quake* discloses that a t-statistic is a statistical method known in the art. Ex. 1008, 5:64-67 ("A commonly used measure of statistical significance when a highly significant result is desired is $p<0.01$, i.e., a 99% confidence interval based on a chi-square or t-test."). In my opinion, a person of ordinary skill in the art at the time of the invention of the '415 patent would have applied conventional statistical analyses, such as a t-test statistic, to the methods disclosed in *Lo II* with a reasonable expectation of success. A person of ordinary skill in the art would have been motivated to use the confidence intervals derived from t statistics when evaluating sequence tag density data to determine the disomy of chromosomes in a mixed sample.

P. *Lo I*, *Shimkets*, *Wang*, and *Hillier* and/or *Smith* Teach Each and Every Feature of Claim 17 of the '415 patent

166. Claim 17 recites "[t]he method of claim 13 wherein said mapping includes mapping sequences with one mismatch."

167. As explained above, *Lo I, Shimkets,* and *Wang* teach all of the features of claim 13. These references are silent as to whether one mismatch is allowed between the sequence tags and the corresponding chromosome portions. A person of ordinary skill in the art at the time of the invention of the '415 patent would have allowed for one mismatch when assigning sequence tags to corresponding chromosome portions. Doing so is merely a known technique to improve similar methods in the same way and yields predictable results.

168. As explained above, it was well known at the time of the invention that single nucleotide polymorphisms exist in human DNA sequences obtained from different individuals. It was also known that sequencing methods were not perfect and that errors can exist in sequence tag information. Consequently, methods of aligning a sequence tag to a reference sequence should account for these nucleotide differences/errors. *Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1006, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1006, page 185, Figure 2. In addition, *Smith*

teaches that allowing mismatches when mapping sequences to a reference sequence can improve the accuracy of the mapping. Ex. 1009, page 4.

169. Based at least on this knowledge, a person of ordinary skill in the art at the time of the invention of the '415 patent would have permitted one mismatch in sequence tags of sufficient length to assign to a chromosome portion when aligning sequence tags obtained by sequencing DNA from a biological sample to corresponding chromosome portions of a reference sequence. A person of ordinary skill in the art would have done so to account for the known existence of polymorphisms and sequence errors, thereby increasing the number of usable sequence tags obtained from a sequencing the DNA in the sample. Furthermore, a person of ordinary skill would have known that allowing one mismatch still permits one to assign the sequence tag to its corresponding chromosome portion. A person of ordinary skill in the art at the time of the invention of the '415 patent would have arrived at the invention of claim 17 based on the teaching of *Lo I*, *Shimkets*, *Wang*, and *Hillier* and/or *Smith*.

VII. Conclusion

170. In summary, it is my opinion that the references I have discussed, either alone or in combination. teach all of the features recited in the claims of the '415 patent.

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

172. In signing this declaration, I understand that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross examination in the case and that cross examination will take place within the United States. If cross examination is required of me, I will appear for cross examination within the United States during the time allotted for cross examination.

Dated: June 26, 2013

By: _____
Stacey Bolk Gabriel

Appendix A

Claim Chart of Claim 7 Based on *Lo II*, *Hillier*, and/or *Smith*

Claim Language	*Lo II*	*Hillier*	*Smith*
Claim 7. The method of claim 3 wherein the step of assigning sequence tags to corresponding chromosome portions allows one mismatch.	*Lo II* discloses all the features of claim 3.	*Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1008, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1008, page 185, Figure 2.	*Smith* discloses "[m]aping longer reads with more mismatches increases accuracy." Ex. 1009, page 4. Specifically, *Smith* discloses mapping selectivity and mapping accuracy of sequence alignment with 0, 1, 2, 3, and 4 mismatches depending on the length of the sequence reads. Ex. 1009, page 4, Figure 2.

Appendix B

Claim Chart of Claims 13 and 16 Based on *Lo II* and *Wang*

Claim Language	*Lo II*	*Wang*
Claim 13. A method of determining an abnormally distributed chromosome portion of interest in a mixed sample of normally and abnormally distributed DNA molecules, comprising:	*Lo II* discloses methods "for determining whether a nucleic acid sequence imbalance (e.g., chromosome imbalance) exists within a biological sample obtained from a pregnant female." Ex. 1002, [0014]. *Lo II* also discloses that the "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." Ex. 1002, [0067]. *Lo II* discloses a "biological sample," which is "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." Ex. 1002, [0033]. "The biological sample may be plasma, urine, serum, or any other suitable sample." Ex.	

Claim Language	*Lo II*	*Wang*
	1002, [0054]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained in the biological sample, and that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054].	
(a) sequencing DNA in said sample by massively parallel sequencing to obtain a number of sequence tags;	*Lo II* discloses that "[a] portion of the nucleic acid molecules contained in the biological sample are sequenced." Ex. 1002, [0015]. *Lo II* also explains that "at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced[,]" and "the nucleic acid molecules are fragments of respective chromosomes." Ex. 1002, [0055]. *Lo II* discloses that the sequencing is done at random. That is, "[t]he origin of a particular fragment is not selected ahead of time." Ex. 1002, [0080]. Because "[t]he sequencing is done at random … a database search may be performed to see where a particular fragment is coming from[,]" indicating that the sequence tag must be of sufficient length to assign the sequence to a location on	

Claim Language	*Lo II*	*Wang*
	a chromosome of the genome. Ex. 1002, [0080]. *Lo II* discloses, as one embodiment, performing the sequencing employed in the aneuploidy detection methods using massively parallel sequencing, which "allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion." Ex. 1002, [0056]. The Illumina Genome Analyzer (or Solexa platform) was identified by *Lo II* as a suitable instrument for performing massively parallel sequencing. *Id.*	
(b) mapping said sequence tags to specific chromosome portions, each chromosomal portion being comprised in a sliding window of a predetermined length;	*Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II* discloses that "[a]fter the massively parallel sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to	*Wang* discloses a digital karyotyping method "that provides quantitative analysis of DNA copy number at high resolution." Ex. 1005, Abstract. The method involves first obtaining short sequence tags (21 bp each) from specific locations in the genome. Ex. 1005, page 16156. "These tags generally contain sufficient information to uniquely identify the genomic loci from which they were derived. Second, populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be

Claim Language	*Lo II*	*Wang*
	see where a particular fragment is coming from." Ex. 1002, [0080].	sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." *Id.* Such a method "can accurately identify regions whose copy number is abnormal." Ex. 1005, page 16161. *Wang* further discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. Depending on the purpose of analysis, e.g., chromosome arms, amplifications, and deletions, the size of the windows can be different, such as about 4 MB, 200 kb, and 600 kb. Ex. 1005, page 16158, Table 1. Tag densities in a test cell can also be normalized to the tag densities of a reference cell in the same sliding windows. Ex. 1005, page 16159, Figure 2.
(c) determining numbers of sequence tags mapped to each sliding window on at least each autosome;	*Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II*	Among other things, *Wang* discloses that "populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital

Claim Language	*Lo II*	*Wang*
	discloses that "[a]fter the massively parallel sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from." Ex. 1002, [0080]. *Lo II* discloses, in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses, in the same context, that particular "chromosomal regions" are distinct from chromosomes: "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	enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." Ex. 1005, page 16156. *Wang* discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 1659, and 16160. *Wang* discloses using a sliding windows analysis in methods of digital karyotyping which can detect, among other things, whole chromosome changes. *Wang* discloses using the method to order sequence tags along each chromosome. *Id.*

Claim Language	*Lo II*	*Wang*
	particular chromosome(s) or chromosomal regions." Ex. 1002, [0060]. *Lo II* discloses using chromosomal regions, or sets of chromosomal regions, to determine if aneuploidy exists: "[t]his determination [of increase or decrease of a clinically-relevant chromosomal region] may be done by using 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. Nucleic acid molecules of the biological sample are sequenced, such that a fraction of the genome is sequenced, and the amount may be determined from results of the sequencing. 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)." Ex. 1002, [0050]. *Lo II* also discloses that	

Claim Language	*Lo II*	*Wang*
	"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." Ex. 1002, [0067]. *Lo II* also discloses random sequencing a representative fraction of DNA molecules in a sample and then analyzing the chromosomal regions (that is, the chromosomal windows) to which they align: "[t]he 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." Ex. 1002, [0108].	
(d) determining a mean of said numbers for each		*Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags

Claim Language	*Lo II*	*Wang*
autosome and a second mean for at least all autosomes;		were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157.
(e) calculating a normalized value from all autosomes, using said second mean; and	*Lo II* discloses normalizing sequence tag density data to account for differences in the relative sizes of chromosomes. Ex. 1002, [0069].	*Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157.
(f) comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest.	*Lo II* discloses using the sequencing results to determine first and second amounts of sequences identified as originating from a first and a second chromosome. From those amounts, "[a] parameter from the first amount and the second amount is then compared to one or more cutoff values. Based on the comparison, a classification of whether a fetal chromosomal aneuploidy exists for the first chromosome is determined." Ex. 1002, [0016]. *Lo II* also states that "the fractional count of the	*Wang* also discloses using normalized sequence tag densities evaluated over sliding windows to detect chromosomal aberrations. Ex. 1005, page 16157, and Fig. 1. In addition, *Wang* discloses a comparison of chromosome number analysis for all 22 human autosomes and also the X and Y chromosome. Ex. 1005, page 16158-16159, Table 2.

Claim Language	*Lo II*	*Wang*
	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." Ex. 1002, [0075]. Figs. 4A and 4B in *Lo II* show data for all 22 autosomes and the X and Y chromosomes.	
Claim 16. The method of claim 13 further comprising the step of calculating a normalized value for chromosome X and, if present, Y.	*Lo II* discloses normalizing sequence tag densities and mapping sequence tags to chromosomes X and Y. Ex. 1002, [0069], Figs. 4A and 4B.	*Wang* teaches using normalized values for X and Y chromosomes. Ex. 1005, page 16158-16159, Table 2.

Appendix C

Claim Chart of Claim 14 Based on *Lo II*, *Shimkets*, and/or *Dohm*

Claim Language	*Lo II*	*Shimkets*	*Dohm*
Claim 14. The method of claim 3 further comprising the step of calculating a relationship between numbers of sequence tags and GC content associated with sequence tags in a given window and correcting for a higher or lower number of reads resulting from a change in GC content.	*Lo II* discloses all of the features of claim 3.	*Shimkets* is directed to sequence-based karyotyping. *Shimkets* discloses that "inherent in the sequencing process itself may be a slight bias in favor of sequences with certain compositional characteristics (such as higher or lower GC content, the percentage of nucleotides in a given stretch that are G or C)." Ex. 1004, [0075]. *Shimkets* teaches that "[t]his bias could be ascertained by calibration experiments and then factored in to subsequent computationally derived reference distributions." Ex. 1004, [0075].	*Dohm* "observe[d] a strong correlation between GC richness and read coverage, with the read density being increased in regions of elevated GC content" for the Solexa sequencing platform. Ex. 1010, e104. "Thus, Solexa-based de novo sequencing as well as re-sequencing activities need to calibrate their sequencing output for achieving accordingly high read coverage of AT-rich regions." Ex. 1010, e105.

Appendix D

Claim Chart of Claim 15 Based on *Lo II* and *Quake*

Claim Language	*Lo II*	*Quake*
Claim 15. The method of claim 3 further comprising the step of calculating a t statistic for each chromosome relative to other chromosomes in the mixed sample, whereby each t statistic indicates a value of a chromosome relative to other chromosomes in a sample, said value being indicative of disomy.	*Lo II* discloses all of the features of claim 3.	*Quake* discloses that "[a] commonly used measure of statistical significance when a highly significant result is desired is $p<0.01$, i.e., a 99% confidence interval based on a chi-square or t-test." Ex. 1008, 5:64-67

Appendix E

Claim Chart of Claim 17 Based on *Lo II*, *Wang*, *Hillier*, and/or *Smith*

Claim Language	*Lo II* and *Wang*	*Hillier*	*Smith*
Claim 17. The method of claim 13 wherein said mapping includes mapping sequences with one mismatch.	*Lo II* and *Wang* disclose all of the features of claim 13. *See* App. B.	*Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1008, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1008, page 185, Figure 2.	*Smith* discloses "[m]aping longer reads with more mismatches increases accuracy." Ex. 1009, page 4. Specifically, *Smith* discloses mapping selectivity and mapping accuracy of sequence alignment with 0, 1, 2, 3, and 4 mismatches depending on the length of the sequence reads. Ex. 1009, page 4, Figure 2.

Appendix F

Claim Chart of Claims 1-6 and 8-12 Based on *Lo II* and *Wang*

Claim Language	*Lo II*	*Wang*
Claim 1. A method of testing for an abnormal distribution of a specified chromosome portion in a mixed sample of normally and abnormally distributed chromosome portions obtained from a subject, comprising:	*Lo II* discloses methods "for determining whether a nucleic acid sequence imbalance (e.g., chromosome imbalance) exists within a biological sample obtained from a pregnant female." Ex. 1002, [0014]. *Lo II* also discloses that the "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." Ex. 1002, [0067]. *Lo II* also discloses a "biological sample," which is "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." Ex. 1002, [0033]. "The biological sample may be plasma, urine, serum, or any other suitable sample." Ex. 1002, [0054]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained	

Claim Language	*Lo II*	*Wang*
	in the biological sample, and that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054].	
(a) sequencing the DNA from the mixed sample to obtain sequences from multiple chromosome portions, wherein said sequences comprise a number of sequence tags of sufficient length of determined sequence to be assigned to a chromosome location with a genome;	*Lo II* discloses that "[a] portion of the nucleic acid molecules contained in the biological sample are sequenced." Ex. 1002, [0015]. *Lo II* also explains that "at least a portion of a plurality of the nucleic acid molecules contained in the biological sample are sequenced[,]" and "the nucleic acid molecules are fragments of respective chromosomes." Ex. 1002, [0055]. *Lo II* discloses that the sequencing is done at random. That is, "[t]he origin of a particular fragment is not selected ahead of time." Ex. 1002, [0080]. Because "[t]he sequencing is done at random … a database search may be performed to see where a particular fragment is coming from[,]" indicating that the sequence tag must be of sufficient length to assign the sequence to a location on chromosome of the genome. Ex. 1002, [0080].	
(b) assigning the sequence tags to corresponding chromosome portions including at least the specified chromosome by	*Lo II* discloses that in its methods "[t]he short sequence tags generated were aligned to the human reference genome sequence and the chromosomal origin was noted." Ex. 1002, [0070]. Similarly, *Lo II* discloses that "[a]fter the massively parallel	

Claim Language	*Lo II*	*Wang*
comparing the determined sequence of the sequence tags to a reference genomic sequence;	sequencing, bioinformatics analysis was performed to locate the chromosomal origin of the sequenced tags." Ex. 1002, [0074]. *Lo II* also discloses that "sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from." Ex. 1002, [0080]	
(c) determining values for numbers of sequence tags mapping to chromosome portions by using a number of windows of defined length within normally and abnormally distributed chromosome portions to obtain a first value and a second value therefrom; and	*Lo II* discloses, in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses, in the same context, that particular "chromosomal regions" are distinct from chromosomes: "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." Ex. 1002, [0060]. *Lo II* discloses using chromosomal regions, or sets of chromosomal regions, to determine if aneuploidy exists: "[t]his determination [of increase or decrease of a clinically-relevant chromosomal region] may	*Wang* discloses a digital karyotyping method "that provides quantitative analysis of DNA copy number at high resolution." Ex. 1005, Abstract. The method involves first obtaining short sequence tags (21 bp each) from specific locations in the genome. Ex. 1005, page 16156. "These tags generally contain sufficient information to uniquely identify the genomic loci from which they were derived." *Id.* "Second, populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag

Claim Language	*Lo II*	*Wang*
	be done by using 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. Nucleic acid molecules of the biological sample are sequenced, such that a fraction of the genome is sequenced, and the amount may be determined from results of the sequencing. 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)." Ex. 1002, [0050]. *Lo II* also discloses that "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." Ex. 1002, [0067]. *Lo II* also discloses random sequencing a representative fraction of DNA molecules in a sample and then analyzing the chromosomal regions (that is, the chromosomal windows) to which they align:	observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution." *Id.* Such a method "can accurately identify regions whose copy number is abnormal." Ex. 1005, page 16161. *Wang* further discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. Depending on the purpose of analysis, e.g., chromosome arms, amplifications, and deletions, the size of the windows can be different, such as about 4 MB, 200 kb, and 600 kb. Ex. 1005, page 16158, Table 1. Tag densities in a test cell can also be normalized to the tag densities of a reference cell in the same sliding windows. Ex. 1005, page 16159, Figure 2.

Claim Language	*Lo II*	*Wang*
	"[t]he 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." Ex. 1002, [0108].	
(d) using the values from step (c) to determine a differential, between the first value and the second value, which is determinative of whether or not the abnormal distribution exists.	*Lo II* discloses using the sequencing results to determine first and second amounts of sequences identified as originating from a first and a second chromosome. From those amounts, "[a] parameter from the first amount and the second amount is then compared to one or more cutoff values. Based on the comparison, a classification of whether a fetal chromosomal aneuploidy exists for the first chromosome is determined." Ex. 1002, [0016].	
Claim 2. The method of claim 1 wherein to determine a differential includes the step of comparing a normalized sequence tag density of the specified DNA chromosome	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses deriving a parameter from a first amount and a second amount: "[b]ased on the	

Claim Language	*Lo II*	*Wang*
portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.	sequencing, a first amount of a first chromosome is determined from sequences identified as originating from the first chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure is found in [0074]. Ex. 1002, [0074]. Fig. 4B in *Lo II* shows data for all 22 autosomes and the X and Y chromosomes.	
Claim 3. The method of claim 1 wherein the mixed sample comprises a mixture of maternal and fetal DNA and wherein the abnormal distribution results from a fetal aneuploidy.	*Lo II* discloses a "biological sample," which is "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." Ex. 1002, [0033]. "The biological sample may be plasma, urine, serum, or any other suitable sample." Ex. 1002, [0054]. *Lo II* further discloses that "nucleic acid molecules from the fetus and the pregnant female" are contained in the biological sample, and that "the nucleic acid molecules may be fragments from chromosomes." Ex. 1002, [0054]. *Lo II* discloses an "invention [that] generally relates to the diagnostic	

Claim Language	*Lo II*	*Wang*
	testing of fetal chromosomal aneuploidy by determining imbalances between different nucleic acid sequences, and more particularly to the identification of trisomy 21 (Down syndrome) and other chromosomal aneuploidies via testing a maternal sample (e.g. blood)." Ex. 1002, [0003]. *Lo II* also discloses that "[f]etal chromosomal aneuploidy results from the presence of abnormal dose(s) of a chromosome or chromosomal region[,]" which may be abnormally high, as in the case of trisomy for chromosome 21. Ex. 1002, [0004]. The abnormal dose(s) can be abnormally high, e.g., the presence of an extra chromosome 21 or chromosomal region in trisomy 21. *Id.*	
Claim 4. The method of claim 1 wherein the mixed sample comprises a mixture of normal and genetically altered DNA from a tumor.	*Lo II* discloses that the clinically relevant chromosomal region and the background nucleic acid may come from first and second cell types. According to *Lo II*, "the percentage of fetal sequences in a sample may be determined by any fetal-derived loci and not limited to measuring the clinically-relevant nucleic acid sequences." Ex. 1002, [0052]. *Lo II* further states that "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	

Claim Language	*Lo II*	*Wang*
	urine, which contains a background of nucleic acid sequences derived from the non-malignant cells within the body." *Id.* *Lo II* also discloses as "clinically relevant nucleic acid sequences" nucleic acid "sequences which are mutated, deleted, or amplified in a malignant tumor, e.g. sequences in which loss of heterozygosity or gene duplication occur." Ex. 1002, [0037].	
Claim 5. The method of claim 3 wherein the sequencing is massively parallel sequencing.	*Lo II* discloses, as one embodiment, performing the sequencing employed in the aneuploidy detection methods using massively parallel sequencing, which "allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion." Ex. 1002, [0056]. The Illumina Genome Analyzer (or Solexa platform) is identified by *Lo II* as a suitable instrument for performing massively parallel sequencing. *Id.*	
Claim 6. The method of claim 3 wherein the fetal aneuploidy is an aneuploidy of at least one of chromosome 13, 18 and 21.	*Lo II* discloses that "a parameter (e.g. a fractional representation) of a chromosome potentially involved in a chromosomal aneuploidy, e.g. chromosome 21 or chromosome 18 or chromosome 13, may then be calculated from the results of the bioinformatics procedure." Ex. 1002, [0063]. Moreover, claim 5 in *Lo II* recites chromosomes 21,	

Claim Language	*Lo II*	*Wang*
	18, and 13 as the chromosomes for which aneuploidy is being tested. Ex. 1002, page 11.	
Claim 8. The method of claim 3 wherein the sequence tags are about 25-100 bp in length.	*Lo II* exemplifies generating sequence tags that are 36 bp in length. Ex. 1002, [0111].	
Claim 9. The method of claim 8 wherein at least about 1 million sequence tags are obtained.	*Lo II* discloses that "[a]s a high number of sequencing reads, in the order of hundred thousands to millions or even possibly hundreds of millions or billions, are generated from each sample in each run, the resultant sequenced reads form a representative profile of the mix of nucleic acid species in the original specimen." Ex. 1002, [0057]. In addition, Figs. 6 and 8 in Lo II identify samples having more than one million sequenced tags. Ex. 1002, Figs. 6 and 8.	
Claim 10. The method of claim 8 further comprising the step of calculating a normalized sequence tag density of the specified DNA chromosome portion and a	*Lo II* discloses that "a proportion of such sequences [referred to in [0067]] would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of	

Claim Language	*Lo II*	*Wang*
normalized sequence tag density of another DNA chromosome portion in said mixed sample.	chromosome 21 compared with the other chromosomes, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allowing the detection of trisomy 21. The degree of change in the normalized frequency will be dependent on the fractional concentration of fetal nucleic acids in the analyzed sample." Ex. 1002, [0069]. *Lo II* also discloses that "[o]ne 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)." Ex. 1002, [0014].	
Claim 11. The method of claim 10 wherein the calculating a differential includes the step of comparing a normalized sequence tag density of the specified DNA	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses deriving a parameter from a first amount and a second amount: "[b]ased on the sequencing, a first	

Claim Language	*Lo II*	*Wang*
chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.	amount of a first chromosome is determined from sequences identified as originating from the first chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure is found in [0074]. Ex. 1002, [0074]. Fig. 4B in Lo II shows data for all 22 autosomes and the X and Y chromosomes.	
Claim 12. The method of claim 11 further comprising the step of measuring over- and under-representation of a chromosome by determining a sequence tag density for each chromosome in the sample, namely chromosomes 1-22, X and also chromosome Y if present.	*Lo II* discloses in the context of sequence data analysis, normalizing the frequency of sequences that are from a chromosome involved in aneuploidy and sequences that are from the other chromosomes. Ex. 1002, [0069]. *Lo II* also discloses deriving a parameter from a first amount and a second amount: "[b]ased on the sequencing, a first amount of a first chromosome is determined from sequences identified as originating from the first chromosome. A second amount of one or more second chromosomes is determined from sequences identified as originating from one of the second chromosomes. A parameter from	

Claim Language	*Lo II*	*Wang*
	the first amount and the second amount is then compared to one or more cutoff values." Ex. 1002, [0016]. Similar disclosure is found in [0074]. Ex. 1002, [0074]. Fig. 4B in *Lo II* shows data for all 22 autosomes and the X and Y chromosomes.	

Appendix G

Claim Chart of Claim 7 Based on *Lo II*, *Wang*, *Hillier*, and/or *Smith*

Claim Language	*Lo II* and *Wang*	*Hillier*	*Smith*
Claim 7. The method of claim 3 wherein the step of assigning sequence tags to corresponding chromosome portions allows one mismatch.	*Lo II* and *Wang* disclose all features of claim 3. *See* App. F.	*Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1008, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1008, page 185, Figure 2.	*Smith* discloses "[m]aping longer reads with more mismatches increases accuracy." Ex. 1009, page 4. Specifically, *Smith* discloses mapping selectivity and mapping accuracy of sequence alignment with 0, 1, 2, 3, and 4 mismatches depending on the length of the sequence reads. Ex. 1009, page 4, Figure 2.

Appendix H

Claim Chart of Claim 14 Based on *Lo II*, *Wang*, *Shimkets*, and/or *Dohm*

Claim Language	*Lo II* and *Wang*	*Shimkets*	*Dohm*
Claim 14. The method of claim 3 further comprising the step of calculating a relationship between numbers of sequence tags and GC content associated with sequence tags in a given window and correcting for a higher or lower number of reads resulting from a change in GC content.	*Lo II* and *Wang* disclose all of the features of claim 3. *See* App. F.	*Shimkets* is directed to sequence-based karyotyping. *Shimkets* discloses that "inherent in the sequencing process itself may be a slight bias in favor of sequences with certain compositional characteristics (such as higher or lower GC content, the percentage of nucleotides in a given stretch that are G or C)." Ex. 1004, [0075]. *Shimkets* teaches that "[t]his bias could be ascertained by calibration experiments and then factored in to subsequent computationally derived reference distributions." Ex. 1004, [0075].	*Dohm* "observe[d] a strong correlation between GC richness and read coverage, with the read density being increased in regions of elevated GC content" for the Solexa sequencing platform. Ex. 1010, e104. "Thus, Solexa-based de novo sequencing as well as re-sequencing activities need to calibrate their sequencing output for achieving accordingly high read coverage of AT-rich regions." Ex. 1010, e105.

Appendix I

Claim Chart of Claim 15 Based on *Lo II, Wang,* and *Quake*

Claim Language	*Lo II and Wang*	*Quake*
Claim 15. The method of claim 3 further comprising the step of calculating a t statistic for each chromosome relative to other chromosomes in the mixed sample, whereby each t statistic indicates a value of a chromosome relative to other chromosomes in a sample, said value being indicative of disomy.	*Lo II* and *Wang* disclose all of the features of claim 3. *See* App. F.	*Quake* discloses that "[a] commonly used measure of statistical significance when a highly significant result is desired is $p<0.01$, i.e., a 99% confidence interval based on a chi-square or t-test." Ex. 1008, 5:64-67

Appendix J

Claim Chart of Claims 1-6 and 8-12 Based on *Lo I* and *Shimkets*

Claim Language	*Lo I*	*Shimkets*
Claim 1. A method of testing for an abnormal distribution of a specified chromosome portion in a mixed sample of normally and abnormally distributed chromosome portions obtained from a subject, comprising:	*Lo I* discloses a method for detecting fetal chromosomal aneuploidies, using the example of trisomy 21, by performing random sequencing of DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192]. The DNA fragments, or genomic sequences, would have originally come from either the fetus or the mother. Ex. 1003, [0192]. In other words, *Lo I* discloses a method for testing for an abnormal distribution of a specified chromosome portion (e.g., chromosome 21) in a mixed sample containing normally and abnormally distributed chromosome portions.	
(a) sequencing the DNA from the mixed sample to obtain sequences from multiple chromosome portions, wherein said sequences comprise a number of sequence tags of sufficient length of determined sequence to be assigned to a chromosome location with a	*Lo I* discloses that one may "do random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother." Ex. 1003, [0192]. According to *Lo I*, "[a] proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be	*Shimkets* discloses a sequence-based karyotyping method that "may be used to determine chromosomal abnormalities including balanced and unbalanced chromosomal rearrangements, polyploidy, aneuploidy, deletions, duplications, copy number polymorphisms and the like." Ex. 1004, [0063]. The method comprises "generating a pool of

Claim Language	*Lo I*	*Shimkets*
genome;	derived from the other chromosomes." Ex. 1003, [0192].	fragments of genomic DNA by a random fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution expected by chance." Ex. 1004, [0007]; Figure 9. The at least 20 contiguous bases obtained "will typically allow the mapping of the fragment to a unique location in a genomic scaffold." Ex. 1004, [0071].
(b) assigning the sequence tags to corresponding chromosome portions including at least the specified chromosome by comparing the determined sequence of the sequence tags to a reference genomic sequence;	*Lo I* discloses a method in which "[t]he general principle … is that if one is to do random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother. A proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be	*Shimkets* discloses "generating a pool of fragments of genomic DNA by a random fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution

Claim Language	*Lo I*	*Shimkets*
	derived from the other chromosomes." Ex. 1003, [0192].	expected by chance." Ex. 1004, [0007]; Figure 9.
(c) determining values for numbers of sequence tags mapping to chromosome portions by using a number of windows of defined length within normally and abnormally distributed chromosome portions to obtain a first value and a second value therefrom; and	*Lo I* discloses normalizing the data obtained from the mapped sequences to account for differences in the respective sizes of different chromosomes. Ex. 1003, [0192] ("By taking into account of the relative size of chromosome 21 compared with the other chromosome, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allow the detection of trisomy 21.").	*Shimkets* discloses normalizing the data obtained from mapped sequences. The "[r]atios, on a per chromosomal basis, of the number of uniquely mapping fragments in the experimental sample to the number in the normal sample (corrected by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples) can be used to estimate rates of aneuploidy." Ex. 1004, [0267]. *Shimkets* also discloses normalizing data by obtaining the distribution of the fragments using a number of windows of defined length within a test chromosome (either normal or abnormal) and a normal chromosome. "The number of a plurality of sequences

Claim Language	*Lo I*	*Shimkets*
		mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, [0007]. *Shimkets* further explains the concept of "windows" in relation to "the test cell distribution (i.e., chromosomal map density): "The test cell distribution (i.e., chromosomal map density) is defined as the number of mapped sequences (i.e., fragments) by the number of possible map locations present in a given chromosome. The number of possible map locations is defined by the

Claim Language	*Lo I*	*Shimkets*
		size of the observation window and the length of the chromosome." Ex. 1004, [0012], [0073].
(d) using the values from step (c) to determine a differential, between the first value and the second value, which is determinative of whether or not the abnormal distribution exists.		*Shimkets* discloses that "[t]he number of a plurality of sequences mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, [0007].
Claim 2. The method of claim 1 wherein to determine a differential includes the step of comparing a normalized sequence tag density of the		*Shimkets* discloses that the ratio, on a per chromosomal basis, of the number of mapped sequences in an experimental sample to the number in the normal sample can be normalized "by the ratio of the total number of uniquely

Claim Language	*Lo I*	*Shimkets*
specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.		mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples." Ex. 1004, [0267]. *Shimkets* discloses that "[c]ounts of the resulting number of unique hits to each chromosome were tabulated for both the test DiFi sample and the reference GM12911 sample. For each chromosome, the ratio of the number of unique hits in the DiFi sample to the corresponding number of hits to the GM12911 sample was computed, providing a raw ratio of measured chromosomal content on a per chromosome basis. The raw ratios were further normalized to account for any difference in the amount of actual sequencing performed for the two samples; specifically, the ratio of the total number of unique hits to the autosomal chromosomes in the DiFi and GM12911 samples

Claim Language	*Lo I*	*Shimkets*
		was used as a multiplicative normalization factor to convert the raw chromosomal content ratios into normalized ratios." Ex. 1004, [0248].
Claim 3. The method of claim 1 wherein the mixed sample comprises a mixture of maternal and fetal DNA and wherein the abnormal distribution results from a fetal aneuploidy.	*Lo I* discloses a method for detecting fetal chromosomal aneuploidies,\ using the example of trisomy 21, by performing random sequencing of DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192]. The DNA fragments, or genomic sequences, would have originally come from either the fetus or the mother. Ex. 1003, [0192].	
Claim 4. The method of claim 1 wherein the mixed sample comprises a mixture of normal and genetically altered DNA from a tumor.	*Lo I* describes methods using a mixed sample comprising a mixture of normal and genetically altered DNA. Ex. 1003, [0192].	*Shimkets* discloses that "Sequence-Based Karyotyping or high resolution molecular karyotyping according to the invention can be used to identify remaining oncogenes and tumor suppressor genes...." Ex. 1004, [0092]. Shimkets discloses this embodiment as a comparison of "the genomes from a normal subject and a diseased subject." *Id.*
Claim 5. The	*Lo I* discloses methods involving	

Claim Language	*Lo I*	*Shimkets*
method of claim 3 wherein the sequencing is massively parallel sequencing.	random sequencing using massively parallel genomic sequencing. Ex. 1003, [0192].	
Claim 6. The method of claim 3 wherein the fetal aneuploidy is an aneuploidy of at least one of chromosome 13, 18 and 21.	*Lo I* discloses a method for detecting fetal chromosomal aneuploidies, using the example of trisomy 21, by performing random sequencing of DNA fragments present in the plasma of a pregnant woman. Ex. 1003, [0192].	
Claim 8. The method of claim 3 wherein the sequence tags are about 25-100 bp in length.		*Shimkets* discloses that "[w]hile the sequencing of 20 bp from each fragment is sufficient, sequencing of more bases is also useful. For example, the sequencing of at least 25 bp, at least 30 bp, at least 35 bp, at least 40 bp, at least 45 bp, at least 50 bp, at least 55 bp, at least 60 bp, at least 65 bp, at least 70 bp, at least 75 bp, at least 80 bp, at least 95 bp, at least 100 bp have been performed by the methods of the invention and found to be useful but not essential." Ex. 1004, [0070].
Claim 9. The method of claim 8		*Shimkets* discloses that "[a]t least 1000, 10,000,

Claim Language	*Lo I*	*Shimkets*
wherein at least about 1 million sequence tags are obtained.		100,000, 1,000,000 or more sequenced are mapped." Ex. 1004, [0011].
Claim 10. The method of claim 8 further comprising the step of calculating a normalized sequence tag density of the specified DNA chromosome portion and a normalized sequence tag density of another DNA chromosome portion in said mixed sample.	*Lo I* teaches using a mixed sample. Ex. 1003, [0192].	*Shimkets* discloses that the ratio, on a per chromosomal basis, of the number of mapped sequences in an experimental sample to the number in the normal sample can be normalized "by the ratio of the total number of uniquely mapping sequences to the entire genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples." Ex. 1004, [0267]. *Shimkets* discloses calculating normalized ratios for the autosomal chromosomes from normal (reference GM12911) and abnormal (DiFi) cells. Ex. 1004, [0248].
Claim 11. The method of claim 10 wherein the calculating a differential includes the step of comparing a		*Shimkets* discloses that "[t]he raw ratios were further normalized to account for any difference in the amount of actual sequencing performed for the two samples;

Claim Language	*Lo I*	*Shimkets*
normalized sequence tag density of the specified DNA chromosome portion to a normalized sequence tag density of another DNA chromosome portion in said mixed sample, wherein all autosomes are used to calculate the normalized sequence tag density.		specifically, the ratio of the total number of unique hits to the autosomal chromosomes in the DiFi and GM12911 samples was used as a multiplicative normalization factor to convert the raw chromosomal content ratios into normalized ratios." Ex. 1004, [0248].
Claim 12. The method of claim 11 further comprising the step of measuring over- and under-representation of a chromosome by determining a sequence tag density for each chromosome in the sample, namely chromosomes 1-22, X and also chromosome Y if present.		*Shimkets* states: "In the extreme, one could make a contingency table of the entire genome, with one column per chromosome to identify chromosomes that are over or underrepresented in content at the entire chromosomal level. Ratios, on a per chromosomal basis, of the number of uniquely mapping fragments in the experimental sample to the number in the normal sample (corrected by the ratio of the total number of uniquely mapping sequences to the entire

Claim Language	*Lo I*	*Shimkets*
		genome of the normal sample over the number in the experimental sample, to correct for differences in the amount of sequencing in the two samples), can be used to estimate rates of aneuploidy." Ex. 1004, [0267].

Appendix K

Claim Chart of Claim 7 Based on *Lo I*, *Shimkets*, *Hillier*, and/or *Smith*

Claim Language	*Lo I and Shimkets*	*Hillier*	*Smith*
Claim 7. The method of claim 3 wherein the step of assigning sequence tags to corresponding chromosome portions allows one mismatch.	*Lo I* and *Shimkets* disclose all of the features of claim 3. *See* App. J.	*Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1008, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1008, page 185, Figure 2.	*Smith* discloses "[m]aping longer reads with more mismatches increases accuracy." Ex. 1009, page 4. Specifically, *Smith* discloses mapping selectivity and mapping accuracy of sequence alignment with 0, 1, 2, 3, and 4 mismatches depending on the length of the sequence reads. Ex. 1009, page 4, Figure 2.

Appendix L

Claim Chart of Claims 13 and 16 Based on *Lo I*, *Shimkets*, and *Wang*

Claim Language	*Lo I*	*Shimkets*	*Wang*
Claim 13. A method of determining an abnormally distributed chromosome portion of interest in a mixed sample of normally and abnormally distributed DNA molecules, comprising:	*Lo I* discloses a method for testing for an abnormal distribution of a specified chromosome portion (e.g., chromosome 21) in a mixed sample containing normally and abnormally distributed chromosome portions. Ex. 1003, [0192].		
(a) sequencing DNA in said sample by massively parallel sequencing to obtain a number of sequence tags;	*Lo I* discloses methods involving random sequencing using massively parallel genomic sequencing to obtain a number of "genomic sequences" (i.e., sequence tags). Ex. 1003, [0192].	*Shimkets* discloses using a massively parallel sequencing platform, a pyrophosphate sequencer from 454 Life Sciences (New Haven, Conn.), which is capable of sequencing 70,000 beads simultaneously. Ex. 1004, [0580].	
(b) mapping said sequence tags to specific chromosome portions, each	*Lo I* discloses a method in which "[t]he general principle … is that if one is to do	*Shimkets* discloses "generating a pool of fragments of genomic DNA by a random	*Wang* discloses a digital karyotyping method "that provides quantitative

Claim Language	*Lo I*	*Shimkets*	*Wang*
chromosomal portion being comprised in a sliding window of a predetermined length;	random sequencing of DNA fragments that are present in the plasma of a pregnant woman, then one would obtain genomic sequences which would originally have come from either the fetus or the mother. A proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes." Ex. 1003, [0192].	fragmentation method, determining the DNA sequence of at least 20 base pairs of each fragment, mapping the fragments to the genomic scaffold of the organism, and comparing the distribution of the fragments relative to a reference genome or relative to the distribution expected by chance." Ex. 1004, [0007]; Figure 9.	analysis of DNA copy number at high resolution." Ex. 1005, Abstract. The method involves first obtaining short sequence tags (21 bp each) from specific locations in the genome. Ex. 1005, page 16156. "These tags generally contain sufficient information to uniquely identify the genomic loci from which they were derived. Second, populations of tags can be directly matched to the assembled genomic sequence, allowing observed tags to be sequentially ordered along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA

Claim Language	*Lo I*	*Shimkets*	*Wang*
			content with high resolution." *Id.* Such a method "can accurately identify regions whose copy number is abnormal." Ex. 1005, page 16161. *Wang* further discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. Depending on the purpose of analysis, e.g., chromosome arms, amplifications, and deletions, the size of the windows can be different, such as about 4 MB, 200 kb, and 600 kb. Ex. 1005, page 16158, Table 1. Tag densities in a test cell can also be normalized to the tag densities of a reference cell in the same sliding windows. Ex. 1005, page 16159, Figure 2.

Claim Language	*Lo I*	*Shimkets*	*Wang*
(c) determining numbers of sequence tags mapped to each sliding window on at least each autosome;		*Shimkets* discloses that "[t]he number of a plurality of sequences mapping within a given window in the population is compared to the number of said plurality of sequences expected to have been sampled within that window or to the number determined to be present in a karyotypically normal genome of the species of the cell. A difference in the number of the plurality of sequences within the window present in the population from the number calculated to be present in the genome of the cell indicates a karyotypic abnormality." Ex. 1004, [0007]. In discussing mapping sequences to chromosomes in the genome, *Shimkets* discloses	*Wang* discloses that tag densities were analyzed along each chromosome by using sliding windows. Ex. 1005, pages 16157, 16159, and 16160. *Wang* discloses using a sliding windows analysis in methods of digital karyotyping which can detect, among other things, whole chromosome changes.

Claim Language	*Lo I*	*Shimkets*	*Wang*
		that "[t]he test cell distribution (i.e., chromosomal map density) is defined as the number of mapped sequences (i.e., fragments) by the number of possible map locations present in a given chromosome. The number of possible map locations is defined by the size of the observation window and the length of the chromosome." Ex. 1004, [0012].	
(d) determining a mean of said numbers for each autosome and a second mean for at least all autosomes;			*Wang* discloses that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157.
(e) calculating a	*Lo I* discloses	*Shimkets* discloses	*Wang* discloses

Claim Language	*Lo I*	*Shimkets*	*Wang*
normalized value from all autosomes, using said second mean; and	normalizing sequence data taking into account the relative sizes of chromosomes. Ex. 1003, [0192].	normalizing the number of sequences mapped to different chromosomal regions. Ex. 1004, [0248], [0267].	that "[t]ag densities for sliding windows containing N virtual tags were determined as the sum of experimental tags divided by the average number of experimental tags in similar sized windows throughout the genome." Ex. 1005, page 16157.
(f) comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest.	*Lo I* discloses comparing normalized values among autosomes to determine any abnormally distributed autosomal chromosome portion of interest. ("If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allow the detection of trisomy 21."). Ex. 1003, [0192].	The entirety of *Shimkets'* sequence-based karyotyping method entails making the comparison recited in clause (f) of claim 13.	*Wang* also discloses using normalized sequence tag densities evaluated over sliding windows to detect chromosomal aberrations. Ex. 1005, page 16157, and Fig. 1. *Wang* discloses a comparison of chromosome number analysis for all 22 human autosomes and also the X and Y chromosome. Ex. 1005, page 16158-59, Table 2.

Claim Language	*Lo I*	*Shimkets*	*Wang*
Claim 16. The method of claim 13 further comprising the step of calculating a normalized value for chromosome X and, if present, Y.	*Lo I* discloses normalizing sequence data taking into account the relative sizes of chromosomes. Ex. 1003, [0192].	*Shimkets* discloses normalizing the number of sequences mapped to different chromosomal regions. Ex. 1004, [0248], [0267].	*Wang* teaches using normalized values for X and Y chromosomes. Ex. 1005, page 16158-59, Table 2.

Appendix M

Claim Chart of Claim 14 Based on *Lo I*, *Shimkets*, and/or *Dohm*

Claim Language	*Lo I*	*Shimkets*	*Dohm*
Claim 14. The method of claim 3 further comprising the step of calculating a relationship between numbers of sequence tags and GC content associated with sequence tags in a given window and correcting for a higher or lower number of reads resulting from a change in GC content.	*Lo I* and *Shimkets* disclose all of the features of claim 3. *See* App. J.	*Shimkets* is directed to sequence-based karyotyping. *Shimkets* discloses that "inherent in the sequencing process itself may be a slight bias in favor of sequences with certain compositional characteristics (such as higher or lower GC content, the percentage of nucleotides in a given stretch that are G or C)." Ex. 1004, [0075]. *Shimkets* teaches that "[t]his bias could be ascertained by calibration experiments and then factored in to subsequent computationally derived reference distributions." Ex. 1004, [0075].	*Dohm* "observe[d] a strong correlation between GC richness and read coverage, with the read density being increased in regions of elevated GC content" for the Solexa sequencing platform. Ex. 1010, e104. "Thus, Solexa-based de novo sequencing as well as re-sequencing activities need to calibrate their sequencing output for achieving accordingly high read coverage of AT-rich regions." Ex. 1010, e105.

Appendix N

Claim Chart of Claim 15 Based on *Lo I*, *Shimkets*, and *Quake*

Claim Language	*Lo I* and *Shimkets*	*Quake*
Claim 15. The method of claim 3 further comprising the step of calculating a t statistic for each chromosome relative to other chromosomes in the mixed sample, whereby each t statistic indicates a value of a chromosome relative to other chromosomes in a sample, said value being indicative of disomy.	*Lo I* and *Shimkets* disclose all of the features of claim 3. *See* App. J.	*Quake* discloses that "[a] commonly used measure of statistical significance when a highly significant result is desired is $p<0.01$, i.e., a 99% confidence interval based on a chi-square or t-test." Ex. 1008, 5:64-67

Appendix O

Claim Chart of Claim 17 Based on *Lo I*, *Shimkets*, *Wang*, *Hillier*, and/or *Smith*

Claim Language	*Lo I*, *Shimkets*, and *Wang*	*Hillier*	*Smith*
Claim 17. The method of claim 13 wherein said mapping includes mapping sequences with one mismatch.	*Lo I*, *Shimkets*, and *Wang* disclose all of the features of claim 13. *See* App. L.	*Hillier* discloses accounting "for mismatches resulting from sequencing errors or polymorphisms." Ex. 1008, page 183. *Hillier* also determined that ~80% of the reads exhibited 0 or 1 mismatch when uniquely aligned to the reference genome. Ex. 1008, page 185, Figure 2.	*Smith* discloses "[m]aping longer reads with more mismatches increases accuracy." Ex. 1009, page 4. Specifically, *Smith* discloses mapping selectivity and mapping accuracy of sequence alignment with 0, 1, 2, 3, and 4 mismatches depending on the length of the sequence reads. Ex. 1009, page 4, Figure 2.