UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ILLUMINA, INC.,
Petitioner,

v.

MOLECULAR LOOP BIOSCIENCES, INC.,
Patent Owner.

Patent No. 11,840,730

IPR2024-01419

DECLARATION OF STACEY GABRIEL, PH. D. REGARDING U.S. PATENT NO. 11,840,730

TABLE OF CONTENTS

I. INTRODUCTION 1
A. Engagement 1
B. Background and Qualifications 1
C. Compensation 3
D. Information Considered 4
II. LEGAL STANDARDS FOR PATENTABILITY 5
A. Priority Date of Patents and Printed Publications 5
B. Anticipation 7
C. Obviousness 7
III. BACKGROUND INFORMATION ABOUT THE '730 PATENT 9
A. The Prosecution History of the '730 Patent 9
B. Technical Field 12
C. Level of Ordinary Skill in the Art 13
D. Priority Date 14
IV. TECHNICAL BACKGROUND 15
A. DNA and Genetic Mutations 15
B. Next-Generation Sequencing 19
1. Library Preparation 20
2. Sequencing Process 23
3. Analysis of Sequence Reads 25
4. Errors and Bias 28
C. Adding Tags to Nucleic Acids 30
V. ANALYSIS OF THE '730 PATENT 33
A. Overview of the '730 Patent 33
B. Construction of Terms Used in the '730 Patent Claims 34

1. "Set of Differentiator Tags" / "Set of Unique Differentiator Tags" 34
2. "Locus of Origin" 37
3. "Consensus Base Call" 40

VI. IDENTIFICATION OF THE PRIOR ART 42

VII. ANALYSIS OF THE PRIOR ART AND '730 CLAIMS 44

A. Chee Anticipates Claims 1-4 and 9 44

1. Overview of Chee 44
2. Chee was Sufficiently Described in the '442 Provisional Application 48
   a) Chee's Claim 1 48
   (1) Preamble 48
   (2) tagging each individual nucleic acid target in a set of nucleic acid targets to create a set of tagged constructs, with: (a) at least one unique identifier oligonucleotide that distinguishes the tagged construct from other tagged constructs in the set, wherein the at least one unique identifier oligonucleotide is not part of the nucleic acid target; and (b) at least one amplification site; 48
   (3) amplifying the set of tagged constructs; and 50
   (4) determining the sequences of the amplified set of tagged constructs comprising the unique identifier oligonucleotides and the nucleic acid targets. 50
   b) Claim 2 51
3. Chee Describes the Elements of Independent Claims 1 and 9 of the '730 Patent 51
   a) Preamble 52
   b) "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin" 54
   c) Claim 1: "introducing a set of differentiator tags wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci

of origin are associated with more than one differentiator tag" 56

d) Claim 9: "introducing a set of unique differentiator tags wherein members of said set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin".....60

e) "amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons".....62

f) "sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence" 63

g) Claim 1: "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count" 65

h) Claim 9: "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule".....69

4. Claim 2 71

5. Claim 3 72

6. Claim 4 73

B. Chee Renders Obvious Claims 1-4, 7, and 9 74

1. Claims 1 and 9.....74

2. Claim 2 77

3. Claim 3 77

4. Claim 4 77

5. Claim 7 78

C. Chee and Gahan Render Obvious Claims 5-6 and 8.....82

1. '730 Patent Claims 5-6 and 8 82

2. Overview of Gahan 83

3. A Person of Ordinary Skill in the Art Would Have Practiced Chee to Analyze Nucleic Acids According to Conventional Techniques 84
4. Claim 5 86
5. Claim 6 87
6. Claim 8 88

D. Bignell and McCloskey Render Obvious Claims 1, 4-6 and 9 89

1. Overview of Bignell 89
2. Overview of McCloskey 92
3. A Person of Ordinary Skill in the Art Would Have Modified Bignell to Incorporate Elements Shown in McCloskey.....96
4. Bignell and McCloskey Describe the Elements of Independent Claims 1 and 9 of the '730 Patent 101
   a) Preamble 101
   b) "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin" 103
   c) Claim 1: "introducing a set of differentiator tags wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag" 105
   d) Claim 9: "introducing a set of unique differentiator tags wherein members of said set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin".....109
   e) "amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons".....110
   f) "sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence" 111

g) Claim 1: "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count" 113

h) Claim 9: "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule".....115

5. Claim 4 116

6. Claim 5 117

7. Claim 6 117

E. Bignell, McCloskey, and Bentley Render Obvious Claim 2 118

1. Overview of Bentley 118

2. A Person of Ordinary Skill in the Art Would Have Configured Bignell and McCloskey to Generate a Consensus Base Call as Taught by Bentley 119

3. Claim 2 123

4. Claim 3 125

F. Bignell, McCloskey, and Gahan Render Obvious Claims 6 and 8 126

1. A Person of Ordinary Skill in the Art Would Have Practiced Bignell and McCloskey to Analyze Nucleic Acids in Blood For Diagnostic Purposes 126

2. Claim 6 127

3. Claim 8 128

G. Bignell, McCloskey, and Chiang Render Obvious Claim 7 129

1. Overview of Chiang 129

2. A Person of Ordinary Skill in the Art Would Have Practiced Bignell and McCloskey to Determine Copy Number Taught by Chiang 130

3. Claim 7 132

I. INTRODUCTION

A. Engagement

1. I have been retained by counsel for Illumina, Inc. as an expert witness in the above-captioned proceeding. I have been asked to provide an opinion regarding the patentability of certain claims in U.S. Patent No. 11,840,730 ("the '730 patent") (EX1001). The face of the patent identifies Molecular Loop Biosciences, Inc. ("Molecular Loop") as the assignee. I have been asked to provide a discussion of the meaning of certain words and phrases in the claims of the '730 patent, to provide a description of the state of the art of the technology described in the '730 patent, to analyze the disclosure of the '730 patent and the applications to which it claims priority, and to analyze various references that I understand are prior art to this patent.

B. Background and Qualifications

2. I received my B.S. in Molecular Biology from Carnegie Mellon University in 1993 and received my Ph.D. in Genetics from Case Western Reserve University in 1998. After completion of my Ph.D., I joined the Whitehead Institute Center for Genome Research of MIT (WICGR) and became involved in efforts to finalize the work of the Human Genome Project, mainly through working on technologies and applications to discover DNA variation in the human genome.

3. My research interest over the past 25 years has been in large-scale genomics research and its application to human medical and population genetics. I

have played a central role in the development of methods for generating and analyzing data using SNP genotyping and next generation sequencing for both cancer and medical genetics studies. My research on the haplotype structure of the human genome provided an important foundation for the International Human HapMap Project. I have directed our center's execution in flagship large-scale projects including the Human HapMap, the 1000 Genomes Project, The Cancer Genome Atlas, The Trans-omics for Precision Medicine (TOPMed) Project, and the All of Us Research Program. I have served as the Chief Genomics Officer of the Broad Institute and have led the Institute's Genomics Platform, a 250-person organization that operates as one of the largest sequencing centers in the world and continually explores, validates, optimizes, and implements new technologies, methods, and analysis tools to meet the needs of the scientific community.

4. My lab has had a particular focus on cancer and diagnostics. We have been a major data producer for The Cancer Genome Atlas (TCGA). The TCGA is funded by the National Cancer Institute and the National Human Genome Research Institute and was initiated in 2005 to characterize over 10,000 tumor samples across 20 types of tumors. This initiative has been one of the largest and most comprehensive research programs for the field of cancer genomics. I have been heavily involved with TCGA since its conception in various ways. Over the past twenty years, I have been a Principle Investigator for TCGA grants, authored at

least TCGA 15 manuscripts, and served on the Executive and Steering committee. In addition, we have conducted research on cell-free tumor DNA sequencing projects. As part of this work, we have helped developed and have implemented methods for gene panel and genome sequencing of cell-free tumor DNA. As of 2023, my lab has produced genomic profiles for over 10,000 cell-free tumor DNA samples.

5. During the Covid-19 pandemic, I led the development and execution of a commercial-scale diagnostic testing and viral surveillance activity that served all of New England and other states, ultimately producing nearly 38 million Covid-19 PCR diagnostic tests and sequencing of over 100,000 covid viral sequences. In recognition, I was named "Bostonian of the Year" in 2020.

6. Currently, I am Executive Vice President, Platforms and Scientific Execution at the Broad Institute of Harvard and MIT, a position I have held since September 2023. I am a co-author on more than 200 publications in the fields of cancer genomics and human genetics, and have been designated as a "Highly Cited Researcher" by Clarivate from 2014-2022.

9. My Curriculum Vitae is attached as Exhibit A.

C. Compensation

7. I am being compensated for my time at the rate of $900 per hour for my work in connection with this matter. I am being reimbursed for reasonable and

customary expenses associated with my work in this investigation. This compensation is not dependent in any way on the contents of this Declaration, the substance of any further opinions or testimony that I may provide, or the ultimate outcome of this matter.

D. Information Considered

8. My opinions are based on my years of education, research, and experience, as well as my investigation and study of relevant materials. In forming my opinions, I have considered the materials I identify in the attached Exhibit List.

9. I may rely upon these materials and/or additional materials to respond to arguments raised by Molecular Loop. I may also consider additional documents and information in forming any necessary opinions—including documents that may not yet have been identified by me or provided to me.

10. My analysis of the materials produced in this investigation is ongoing, and I will continue to review any new material as it is provided. This report represents only those opinions I have formed to date. I reserve the right to revise, supplement, and/or amend my opinions stated herein based on new information and on my continuing analysis of the materials already provided.

II. LEGAL STANDARDS FOR PATENTABILITY

11. Certain basic legal principles have been explained to me by counsel for Illumina. Below, I have recorded these legal standards as they were explained to me.

12. I understand that for an invention claimed in a patent to be found patentable, it must be, among other things, new and not obvious from what was known before the invention was made. I understand the information that is used to evaluate whether an invention is new and not obvious is generally referred to as "prior art" and can include patents and printed publications.

13. I understand that in this proceeding Illumina has the burden of proving that the challenged claims are unpatentable over the prior art by a preponderance of the evidence. I understand that "a preponderance of the evidence" is evidence sufficient to show that a fact is more likely true than it is not.

14. I understand that there are two ways in which prior art may render a patent claim unpatentable. First, the prior art can be shown to "anticipate" the claim. Second, the prior art can be shown to have made the claim "obvious" to a person of ordinary skill in the art.

A. Priority Date of Patents and Printed Publications

15. I understand that a patent or published patent application will be prior art if its effective filing date is before the earliest effective filing date of the

claimed invention. A printed publication will be prior art if it was publicly available before the earliest effective filing date of the claimed invention. I understand that in this proceeding, the information that may be evaluated to show unpatentability is limited to patents and printed publications.

16. I understand that the effective filing date of a claim in a patent or published patent application is the actual filing date of the application, unless the applicant claims priority to an earlier filed application that supports the claimed subject matter in the manner required by 35 U.S.C. § 112. In that case, the date of the earlier filed application is the effective filing date.

17. I understand that Section 112 requires the patent's specification to contain a sufficient written description of the claimed invention to demonstrate that the applicant actually possessed the invention as of the filing date. In considering whether this written description requirement is met, I understand that I should consider the written description from the viewpoint of a person of ordinary skill in the art. I also understand that I should consider whether this person of ordinary skill would have recognized that the written description describes the full scope of the claimed invention and that the inventor actually possessed that full scope as of the claimed effective filing date.

B. Anticipation

18. I understand that, for a patent claim to be "anticipated" by the prior art, each and every requirement of the claim must be found, expressly or inherently, in a single prior art reference as recited in the claim.

19. I understand that claim limitations that are not expressly described in a prior art reference may still be there if they are "inherent" to the thing or process being described in the prior art.

20. I understand that if a reference incorporates other documents by reference, the incorporating reference and the incorporated reference(s) should be treated as a single prior art reference for purposes of analyzing anticipation.

21. I understand that to be anticipatory, a reference must not only explicitly or inherently disclose every claimed feature, but those features must also be "arranged as in the claim." Differences between the prior art reference and a claimed invention invoke the question of obviousness, not anticipation.

C. Obviousness

22. I understand that a claimed invention is not patentable if it would have been obvious to a person of ordinary skill in the field of the invention at the time the invention was made. I understand that in determining whether a patent claim is obvious, one must consider the following four factors: (i) the scope and content of the prior art, (ii) the differences between the prior art and the claims at issue, (iii)

the knowledge of a person of ordinary skill in the pertinent art; and (iv) objective factors indicating obviousness or non-obviousness, if present (such as commercial success or industry praise).

23. I understand the objective factors indicating obviousness or non-obviousness may include: commercial success of products covered by the patent claims; a long-felt need for the invention; failed attempts by others to make the invention; copying of the invention by others in the field; unexpected results achieved by the invention; praise of the invention by those in the field; the taking of licenses under the patent by others; expressions of surprise by experts and those skilled in the art at the making of the invention; and the patentee proceeded contrary to the accepted wisdom of the prior art. I also understand that any of this evidence must be specifically connected to the invention rather than being associated with the prior art or with marketing or other efforts to promote an invention.

24. I understand that that a patent challenger needs to articulate reasoning for combining known elements. Such reasoning can be based on, for example, design considerations, market demands, looking to solutions to related problems in related fields, and on the "ordinary innovation" and creativity that would be applied by a person of ordinary skill in the art.

25. In addition, I understand that the obviousness inquiry should not be done in hindsight, but must be done using the perspective of a person of ordinary skill in the relevant art as of the effective filing date of the patent claim.

III. BACKGROUND INFORMATION ABOUT THE '730 PATENT

A. The Prosecution History of the '730 Patent

26. I have reviewed the file history of the '730 patent (EX1002-EX1004). I understand that there have been three applications which have been prosecuted: the first application, 13/266,862 ("'862 application"), was filed in March 2012; the second application, 15/231,687 ("'687 application"), was filed in August 2016; the third application, 16/952,764 ("'764" application), was filed in November 2019. The '730 patent issued from the third application.

27. During the prosecution of the first application, 13/266,862, the Applicant focused on claims directed toward molecular inversion probes (MIPs) for over 10 years. *See, e.g.*, EX1002, 122-28, 643-49, 1042-49, 1654-55, 1704-05, 1922-23, 2041-42. In June 2024, the Applicant amended the claims to require that each MIP has a "unique differentiator tag sequence," and that is the first time that element has appeared in the '862 Application. *Id.*, 2152. The application is still being examined. There has not been an office action since the Applicant added that limitation to the '862 Application.

28. The ’687 application was filed as a continuation in 2016, and was initially directed to using unique differentiator tags. EX1003, 33-34. There were numerous office actions where the Examiner rejected the pending claims, followed by numerous amendments and arguments in response by the Application. The main issues had to do with what it meant for a tag to be unique, and whether that was taught in the prior art. *See, e.g.*, *id.*, 428-45, 516-37, 859-77, 985-1005, 1124-53, 1219-48. For example, the Examiner found the term “*unique differentiator tag*” indefinite because it was unclear what the tag had to be unique relative to. *Id.*, 428-30; *also id.*, 516-18, 1124-26. The Examiner also found that the claims were unpatentable based on several prior art references directed to various techniques for tagging nucleic acids (*e.g.*, using sample identifiers or gene-specific identifiers). *Id.*, 430-45; *also id.*, 518-37, 859-77, 985-1005, 1127-53, 1219-48.

29. In response to the rejections, the Applicant’s arguments focused on what the term “*differentiator tags*” meant. The Applicant argued that the invention required each nucleic acid molecule to be tagged with a unique tag such that no other nucleic acid molecule had the same tag. *See, e.g., id.*, 1177. For example, the Applicant stated that “the present claims require ***a unique barcode sequence per each individual target nucleic acid molecule***” when differentiating prior art where a tag with a particular identifier was attached to multiple target molecules. *Id.*, 1177; *see, e.g.*, *id.*, 1177-78 (Applicant arguing art taught associating “the

same barcode" with multiple target molecules, associating "the same barcode for each fragment coming from the same library," or using tags to identify samples or cell types). The inventor, Dr. Porreca, also filed a declaration to support these arguments. *Id.*, 1165-68.

30. In October 2020, the Applicant amended the pending claims of the '687 application to include a step for "correcting" errors or bias for the first time. EX1003, 1258. That language ultimately issued as the claims of the '730 patent. Two of the elements the Applicant added were: (i) "introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality [of target nucleic acid molecules from more than one locus of origin], such that one or more of said loci of origin are associated with more than one differentiator tag;" and (ii) to further require the method to include the step of "correcting for errors or bias. . . by collapsing target:differentiator tag combinations observed more than once into a single count." EX1003, 1258. The Examiner, in response to this amendment, indicated in an interview "it was improper to switch inventions" and that the "originally filed claims did not provide for such limitations and thus the instant claims were a distinct invention." *Id.*, 1274.

31. In November 2020, the Applicant filed the '764 application with claims that included a "correcting" step. EX1004, 356-58. The Examiner,

however, still rejected the claims for being obvious over three prior art references. *Id.*, 866-72; *see also id.*, 994-1002. Additionally, the Examiner found that the specification did not teach the "correcting" step and therefore rejected the claims for lack of written description. *Id.*, 861-62; *see also id.*, 987-91.

32. On March 4, 2021, the Applicant submitted another declaration from Dr. Porreca. EX1004, 964-67. In the declaration, Dr. Porreca stated that none of the prior art disclosed that "***each individual nucleic acid molecule is tagged with a unique differentiator tag***," and that such tags were a "***distinguishing feature of [his] claims***." *Id.*, 965. The Examiner was not persuaded by Applicant's arguments and issued a final rejection. *Id.*, 982-1003.

33. The Applicant appealed to the PTO Board. *See, e.g.*, EX1004, 1071 (citing Porreca Decl. ¶¶6-12 (EX1004, 965-66)), 1057; *also id.*, 1146. After briefing and argument, the Board reversed the Examiner's rejection. The Board found that the claims were not obvious over the prior art references and that the specification disclosed the "correcting" step. EX1004, 1175-82, 1189-92. Following the Board's decision, the Examiner allowed the claims and the '730 patent issued. *Id.*, 1195.

B. Technical Field

34. The '730 patent generally relates to a method of nucleic acid sequencing which uses tags or barcodes to correct for errors or bias.

C. Level of Ordinary Skill in the Art

35. I understand that the claims of a patent are to be reviewed from the point of view of a hypothetical person of ordinary skill in the art at the time of the filing of the patent.

36. I understand that Illumina has taken the position that a person of ordinary skill in the art ("POSA" or "skilled person") at the relevant time frame would have been a person with a (i) Ph.D. in molecular biology, genetics, bioinformatics, or a related field, and at least two years of experience in high-throughput sequencing technologies or (ii) Master's degree in one of the same fields with at least four years of experience in high-throughput sequencing technologies. Based on my knowledge and experience, I agree with the definition proposed by Illumina. I further note that the description of a POSA is approximate, and that varying combinations of education and practical experience also could be sufficient.

37. Well before May 2009, my level of skill in the art was at least that of a person of ordinary skill in the art in the field of the '730 patent, as discussed above. I am qualified to provide opinions concerning what a person of ordinary skill in the art would have known and understood at that time. In preparing this declaration, I have considered the issues from the perspective of a hypothetical person of ordinary skill in the art at the relevant timeframe.

D. Priority Date

38. I understand that the ’730 patent claims priority back to four provisional applications. *See* EX1006; EX1007; EX1008; EX1009.

39. The first provisional application, Provisional Application No. 61/174,470 (“’470 provisional”), was filed on April 30, 2009. *See generally* EX1006. The second provisional application, Provisional Application No. 61/178,923 (“’923 provisional”), was filed May 15, 2009. *See generally* EX1007. Neither the ’470 provisional or the ’923 provisional disclose differentiator tags or unique identifiers. The ’470 provisional also does not disclose correcting for errors or bias. Further, neither the ’470 provisional or the ’923 provisional disclose correcting for errors or bias by “*collapsing target:differentiator tag combinations… into a single count*,” as required by claims 1-8, or “*recognizing target/differentiator tags having the same sequence as being derived from the same input molecule*,” as required by claim 9. Accordingly, I do not believe that the ’470 provisional or the ’923 provisional adequately support the subject matter that is claimed by the ’730 patent.

40. The third provisional application, Provisional Application No. 61/179,358 (“’358 provisional”), was filed May 18, 2009. *See* EX1008. As the earlier provisional applications do not support the claimed subject matter, I understand that the filing date of the ’358 provisional is the earliest date to which

the ’730 patent could potentially claim priority. Therefore, I have employed this priority date in my analysis.

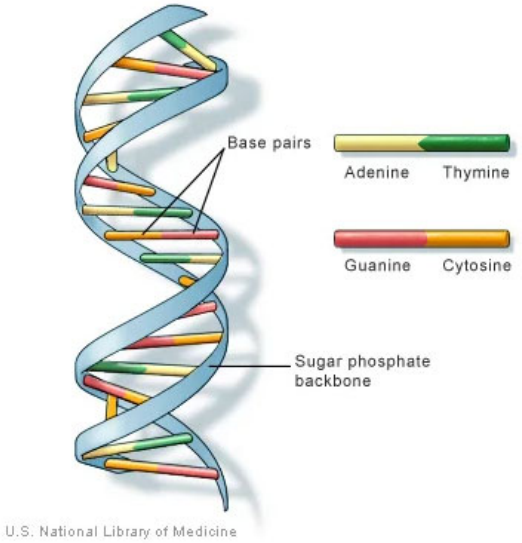
IV. TECHNICAL BACKGROUND

A. DNA and Genetic Mutations

41. Deoxyribonucleic acid (DNA) is a molecule which carries genetic information and instructions for the development and function of organisms. It is often referred to as the blueprint for life.

42. DNA is made of building blocks called nucleotides. There are four different nucleotides for DNA: adenine (A), thymine (T), guanine (G), and cytosine (C). Ribonucleic acid (RNA) has the same bases except that thymine is replaced with uracil (U). The order (*i.e.*, sequence) of these nucleotides conveys information to the organism. This information can provide instructions on how to create certain proteins or how cells should operate.

43. In an organism, each DNA molecule includes two strands of DNA that are bonded to each other. One DNA strand is comprised of numerous nucleotides. The two strands are held together by bonds between the nucleotides. For instance, A binds with T and G binds with C. These units (A-T and G-C) are referred to as “base pairs.” Each strand is complementary to the other. The two strands twist together, giving DNA its double helical structure.



44. The backbone of a DNA molecule is made up of sugar phosphates. The DNA molecule has a directionality or orientation determined by which carbon atom of the sugar-ring is bonded to the phosphate atom in the backbone. One end of the DNA molecule is referred to as the 3' (three prime) end, the other is the 5' (five prime) end. When a strand of DNA is copied, it is synthesized in the 5' to 3' direction.

45. A segment of DNA can be considered a gene. A gene may encode instructions for, for example, making a particular protein. The length of human genes can vary and can be comprised of a couple hundred to more than two million base pairs.

46. The human genome is made up of 23 pairs of chromosomes. For each pair of chromosomes, one comes from the mother and one from the father. Each chromosome includes many genes. The location of a gene can be referred to as a

locus, and it can be identified using a numbering system. For example, the primary genes relating to eye color are located near 15q13.1, which is the 15th chromosome, the long arm (q), region 1, band 3, sub-band 1.

47. By sequencing an organism's DNA, genetic traits or abnormalities can be identified. Some abnormalities are known to be associated with disease.

48. Each cell of an organism includes a complete copy of its DNA. Cellular DNA can be extracted from a tissue sample obtained, for instance, in a tissue biopsy.

49. DNA can also be found circulating in the bloodstream. DNA in the bloodstream is referred to as "cell-free DNA" or "circulating DNA." As cells die naturally in the body, their DNA can be released into the bloodstream. Tumor cells also shed DNA into the bloodstream which can be referred to as circulating tumor DNA. Additionally, fetal cells can shed DNA into the maternal bloodstream which can be referred to as maternally-circulating fetal DNA or maternally-circulating placental DNA.

50. When cells replicate, the DNA in the cell is also replicated in a process called DNA synthesis. The two strands of each DNA molecule separate and nucleotides are added to each strand to synthesize new complementary strands. If there is an error during DNA synthesis, this may cause a genetic mutation. One

type of genetic mutation is a single nucleotide variant where one nucleotide was substituted incorrectly. This is the most common type of genetic mutation.

51. Other examples of genetic mutations include instances where a nucleotide is deleted, or conversely where a nucleotide is added. Deletion and insertion mutations can affect more than one nucleotide. A large segment of a chromosome with multiple genes may be affected. Shorter deletion and insertion mutations which affect less than 50 nucleotides are referred to as "indels." Larger events affecting more than 50 nucleotides are generally referred to as "structural variants."

52. The effect of a genetic mutation depends on the type of mutation as well as the gene that has been mutated. Many mutations may have no consequence on the function of the organism. For instance, a single nucleotide substitution may not impact the coding of the protein. Mutations may also occur in a non-coding region of the genome, which is a region that does not code for proteins. On the other hand, mutations or genetic variations can contribute to severe diseases and genetic illnesses. Most disease-causing variations are "coding" and occur within the genetic portion of the genome.

53. Another type of genetic variation is "copy number variation" or "CNV." The genome includes segments of DNA which repeat a certain number of times (*i.e.*, it includes multiple "copies" of the segment). CNV refers to when

there are an abnormal number of copies of the segment. One example of CNV is aneuploidy where there are an abnormal number of chromosomes. Trisomy 21, for instance, is one example of aneuploidy in which there are three copies of chromosome 21 instead of two. Trisomy 21 is commonly referred to as Down Syndrome.

54. CNV is also commonly associated with cancer. Gains and losses of DNA are prevalent in cancer and are a consequence of various molecular processes that lead to chromosomal instability and ultimately contribute to tumorigenesis by affecting the regulation and/or dosage of genes of multiple genes at the same time. EX1033(Chiang), 99.

B. Next-Generation Sequencing

55. In the mid-2000s, the industry experienced a dramatic shift when several new techniques for DNA sequencing were developed. These "next generation" sequencing methods quickly gained popularity as an alternative to traditional Sanger-based sequencing. "Next-Gen" technologies, also called "Massively Parallel Sequencing," made it possible to produce millions of sequencing reads in a single instrument run (compared to hundreds with "first generation" approaches). It also allowed multiple DNA samples to be sequenced in a single run, which is a process referred to as "multiplex sequencing." Because of this advancement, Next-Gen technologies were widely adopted by 2009. They

had already reduced the cost of sequencing by over two orders of magnitude and had greatly improved efficiency.

56. The power and availability of these Next-Gen platforms led to a reimagining of what could be possible in extracting biological findings and clinical insights through generation of large amounts of sequence data. By May 2009, there were two major platforms that used the sequencing by synthesis method, including Illumina and Roche 454 sequencing.

57. Since the mid-2000s, Next-Gen sequencing platforms have employed diverse sequencing biochemistry, but they generally follow a similar workflow of preparing a library of nucleic acids for sequencing, sequencing the nucleic acids, and analyzing the sequence reads.

1. Library Preparation

58. Library preparation refers to the process of preparing nucleic acids for sequencing. Library preparation is generally accomplished by the fragmentation of DNA, followed by ligation of appropriate adaptor sequences to one or both ends of the DNA fragments, and finally, amplification of the fragments to generate copies. The specification of the ’730 patent acknowledges that there were several known methods of library preparation. *See, e.g.*, EX1001, 5:53-6:37.

59. The DNA is first fragmented because Next-Gen techniques typically operate on shorter segments of DNA than Sanger-based sequencing. Sanger-based

sequencing could read segments of DNA of 500 to 800 nucleotides. Depending on the platform, the early versions of Next-Gen techniques could generate sequence reads that are 50 to 150 nucleotides (nts) or base pairs (bp) in length. Computer algorithms are used to align and assemble the shorter reads to reconstruct the DNA sequence.

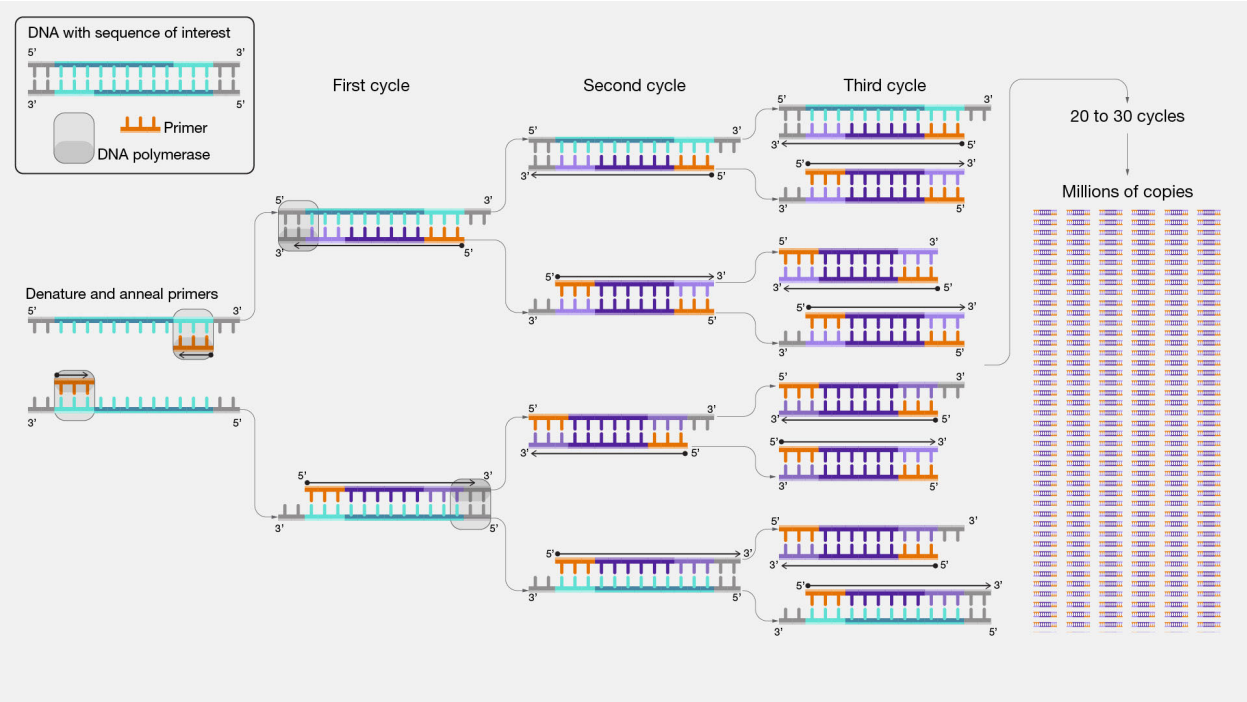
60. Following fragmentation, adaptors are added to the DNA fragments via a ligation reaction. An adaptor sequence is an oligonucleotide (artificially synthesized DNA molecule) attached to a nucleic acid molecule to make manipulation of the molecule easier (*e.g.*, easier to attach the molecule to a surface or to create a ring).

61. The companies that made Next-Gen sequencers typically sold kits that scientists could use to add the appropriate adapters and tags to prepare a sample for sequencing. These kits also were manufactured and sold by third parties. It was within the skill of a POSA to perform the sample preparation, and to make modifications to the sample preparation process to fit the needs or goals of the desired sequencing analysis.

62. The nucleic acid molecules are then subjected to multiple rounds of amplification to generate enough material for sequencing. Amplification is a process for making copies of the original nucleic acid molecules in a sample. Library preparation typically includes one or more amplification steps. One

common, well-known amplification technique was polymerase-chain reaction or PCR. The '730 patent also acknowledges that PCR was a known method of amplification. *See, e.g.*, EX1001, 70:55.

63. PCR creates copies of DNA fragments exponentially as illustrated below:



EX1034, 1.

64. During amplification, a double stranded DNA molecule is first denatured resulting in two individual DNA strands. Short oligonucleotides called primers are added that bind to the individual strands and provide a starting point for the synthesis of a complementary strand. An enzyme called DNA polymerase is added that helps to add nucleotides to the primer and synthesize a

complementary DNA strand. This process turns one double-stranded DNA molecule into two double-stranded DNA molecules. The process is repeated many times, creating many copies. Amplified nucleic acids are often referred to as amplicons.

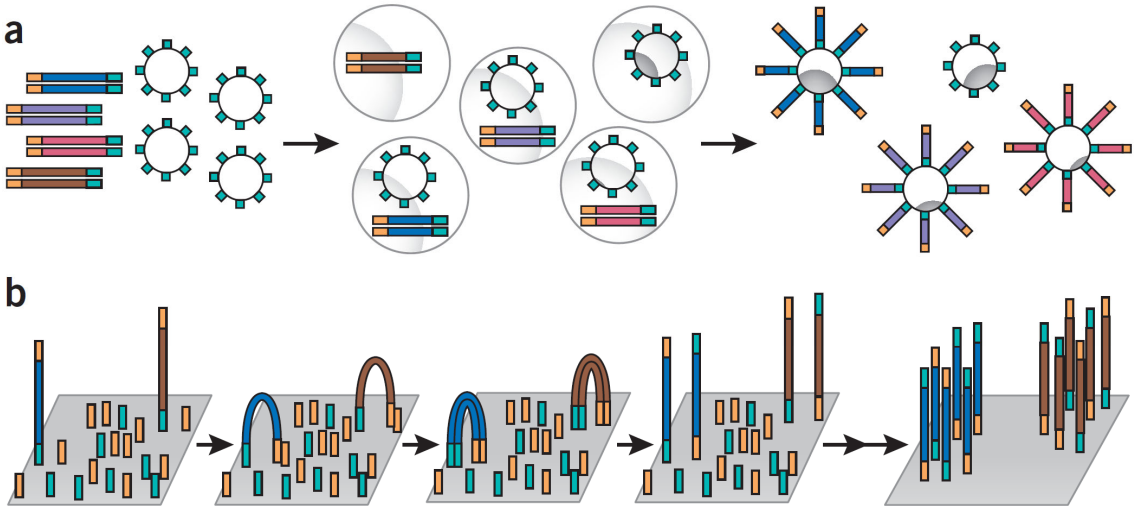
2. Sequencing Process

65. After the library is prepared, the nucleic acid molecules are sequenced. As I explained above, there were several known methods to sequence nucleic acids in May 2009, including the Illumina and Roche 454 Next-Gen sequencing platforms. The specification of the '730 patent also identifies Illumina's sequencing platform as one which was known in the art. *See, e.g.*, EX1001, 72:45-46.

66. Next-Gen platforms used a technique called sequencing-by-synthesis. EX1017(Bentley), 546-48. That process allowed Next-Gen systems to read the sequence of single-stranded template DNA molecules by synthesizing a strand of DNA complementary to each template.

67. After the library was prepared it would be added to a sequencing machine, for example, Illumina's Genome Analyzer (which incorporated Solexa technology). The template DNA molecules in the sample would be spread out inside of the machine. Illumina's platform used flow cells while the 454 platform used beads, but the processes were analogous. For each template DNA molecule,

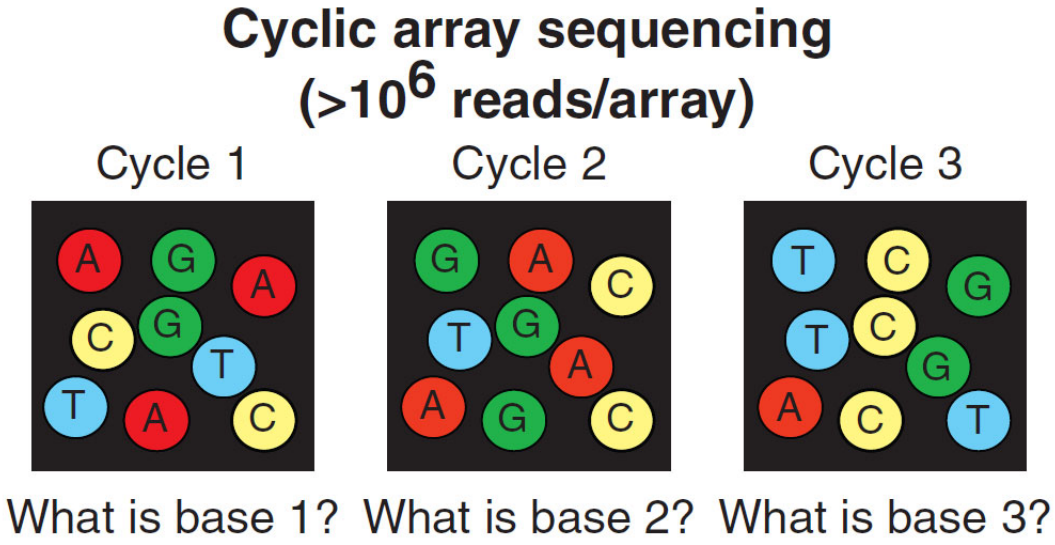
the machine would generate colonies containing multiple copies of that molecule. The clonal amplification steps of the 454 and Illumina platforms can be seen in the figure below as (a) and (b), respectively.



EX1014, 4.

68. Following the clonal amplification step, nucleotides are "read out" as they are incorporated. The 454 platform made use of a single fluorescent indicator, while Illumina made use of nucleotides (A, C, G, and T) that each fluoresce a different color. Nucleotides are added to the complementary strand one-by-one. Each time a nucleotide is added, a laser or other light source is pulsed to read the color of the fluorescent molecule, revealing the nucleotide that was added. This process allows for the determination of the sequence of the synthesized complementary strand, and therefore, of the template strand as well. EX1014, 1139.

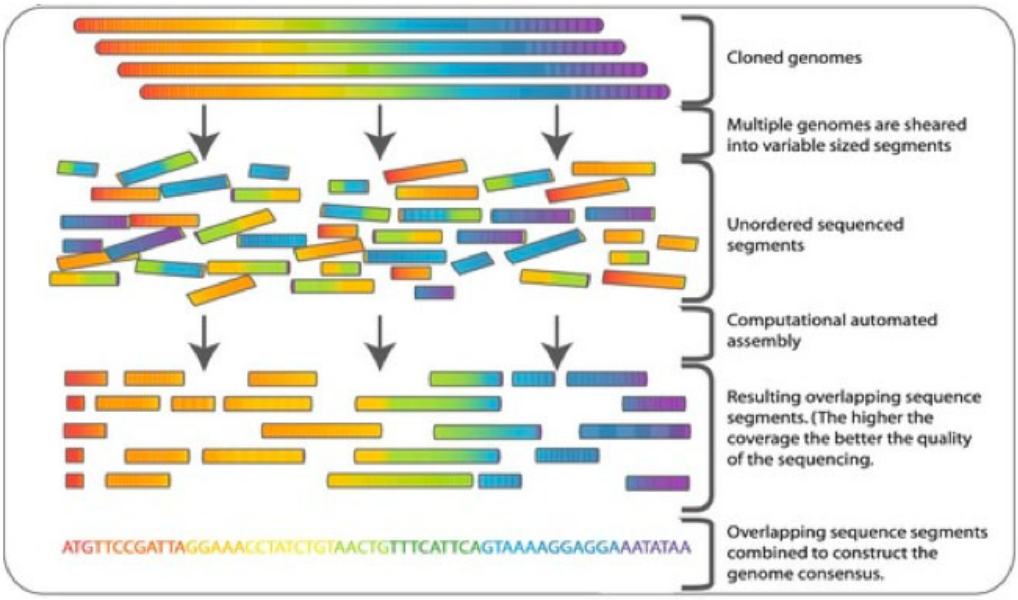
69. The sequencing process is illustrated in the figure below. The black square is a flat surface, and a different nucleic acid is attached to the surface under each circle. At each cycle, one nucleotide is added. A laser is pulsed and the color of the light is measured. The system reads the color of the light and determines the base that was added.



EX1014, 1136. In the figure above, the top left nucleic acid has the sequence A-G-T, while the bottom left nucleic acid has the sequence T-A-A.

3. Analysis of Sequence Reads

70. After the sequence of each fragment is read, computer algorithms are used to align and assemble the reads to reconstruct the original DNA sequence. An illustration of the reconstruction process is below.

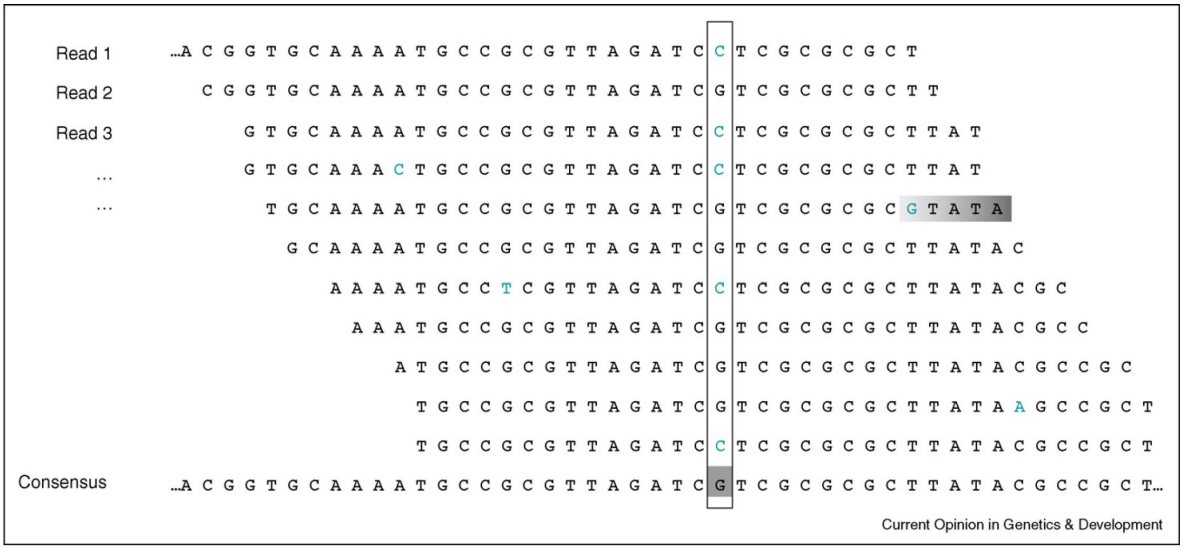


Commins J. et al, Biol Proced Online 11(1) 2015

EX1035, 4.

71. Sequence reads can be compared to a "reference sequence" to identify mutations. Reference sequences can be generated from DNA taken from a healthy person or be a composite of DNA from multiple people. For instance, if the sequence read does not match the reference sequence, that may indicate the presence of a mutation (it could also indicate the presence of an amplification or sequencing error).

72. Multiple sequence reads may also be used to generate a consensus sequence for a particular sample of DNA. The consensus sequence is derived by by determining which nucleotide occurs most frequently at each position in the sequence. At each position, there can be a consensus base call. This is illustrated in the figure below.



Viewing sequence alignments. A schematic diagram of the viewing alignments of 35-base reads derived from a diploid DNA sample, and the resulting consensus, in the style of the gap4 program for sequence alignment and viewing [23] (http://staden.sourceforge.net). The gap4 viewer contains extensive editing and viewing features, including the ability to view quality scores, traces and sequence differences. The schematic shows errors that are distributed randomly (one each of C, T, G and A in blue) and excluded from being true base-calls because they are isolated examples of a base call, different from all the other calls at that position in other reads that contribute to the consensus. In one position (boxed), there are multiple incidences of two base-calls. The consensus quality value is low (grey shading) because of high representation of two base-calls, and this indicates the presence of a heterozygous position, with both base-calls being true. In addition to the use of majority voting and the consensus quality value to make the call, and depending on the sequencing chemistry used, there may be a quality value for each base of each read (see text).

EX1017(Bentley), 550.

73. Sequence reads may also be analyzed to detect copy number. As I explained above, copy number refers to segments of the genome which repeat a certain number of times. Copy number can be analyzed by comparing the number of sequence reads which map to a region of the genome to the number expected to map in a "normal" genome. If the copy number differs from the "euploid" expectation genome, that is indicative of copy number variation or "CNV."

74. Finally, the analysis of the sequence reads may involve comparing mutation frequencies, copy number, or total amount of sequence reads from a sample to a reference sample (also called a control sample). The reference sample

may be from a healthy patient or a patient with a particular disease. In both scenarios, comparing a sample to a reference sample can facilitate diagnosing disease.

4. Errors and Bias

75. It was well-known that errors could be introduced in the amplification and sequencing steps. Amplification copies the sequence by generating a strand that is complimentary to the sequence. In this process, a nucleotide may be incorrectly incorporated in the complementary strand. For instance, if the nucleotide in the original strand was a "T," then an "A" should be added to the complementary strand. If a "C" was added to the complementary strand instead, then the sequence read would show that a "G" was in the position where the "T" should be. This would be considered an amplification error.

76. Further, the sequencing step can also introduce error. Sequencing by synthesis is a sequencing technique which, similar to amplification, occurs by adding nucleotides to a complementary strand. Thus, a nucleotide may also be incorrectly incorporated in the complementary strand in the sequencing step. Alternatively, the nucleotides may be correctly incorporated in the complementary strand, but mis-read by the sequencing machine. This would also be considered a sequencing error.

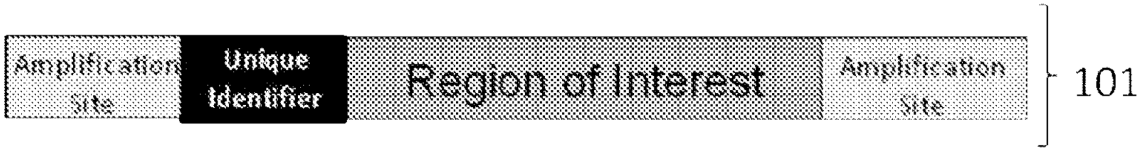
77. The amplification step can also introduce bias. Amplification occurs at random so not every nucleic acid molecule is amplified the same number of times. Some molecules may be over- or under-amplified. For example, an original sample may have one Fragment A and one Fragment B. After amplifying the fragments, there may be 8 copies of Fragment A and only 5 copies of Fragment B. Fragment A was over-amplified. Here, counting the copies of the fragments would not accurately capture the nucleic acid molecules that were originally in the sample. As a result, the sequencing reads will be skewed. This is commonly referred to as amplification bias. This concept is described in many articles from before May 2009. EX1025, 8966 (noting the presence of "skewing" in studies "due to unequal amplification (PCR bias)"); EX1018, [0002], [0089]; EX1010(Chee), 10:11-14; EX1012(McCloskey), [0005] ("skewing [] can arise from PCR amplification"), [0003]; EX1018(Lexow), [0002] ("in an amplification process one target molecule in a sample may be copied more times than another target molecule").

78. Identifying true, biologic genetic variations is more difficult when there are errors or bias. For example, it can be difficult to distinguish an error from a point mutation. It also can be difficult to count the number of copies of a sequence when there is amplificaiton bias. Various techniques had been developed by 2009 to address errors and bias.

C. Adding Tags to Nucleic Acids

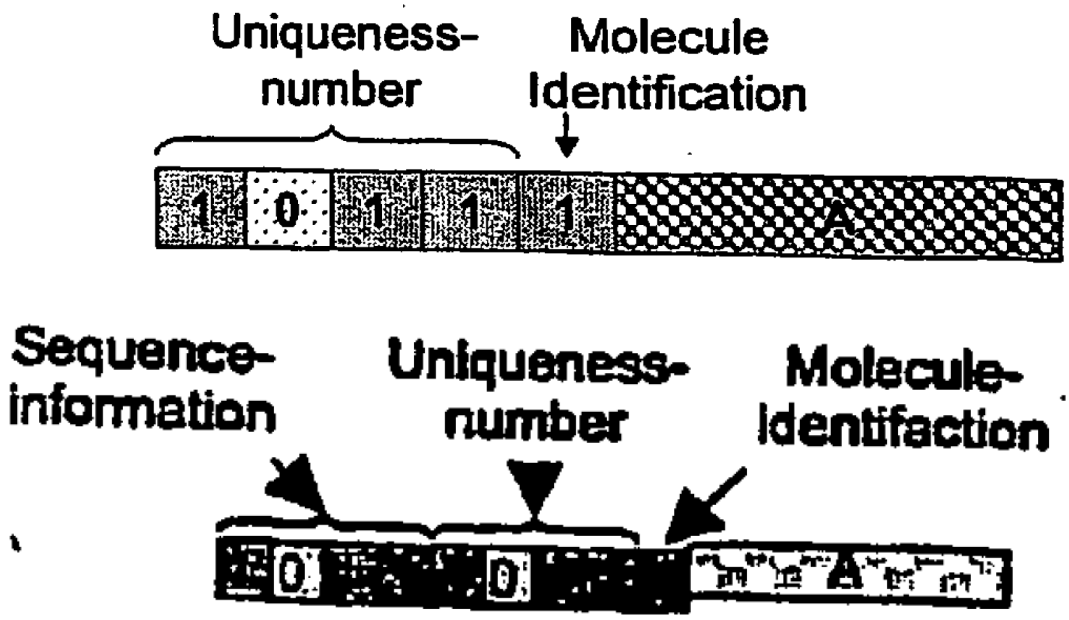
79. By the early 2000s, it was well-known that nucleic acids could be labeled with a oligonucleotide "tag" that included a sequence encoding information about the nucleic acid. Tags were also referred to as "barcodes" or "identifiers."

80. Tags could be ligated onto the end of a nucleic acid like adaptors. Chee shows an example of a nucleic acid molecule (labeled region of interest) with a tag (labeled unique identifier) and two adaptors (labeled amplification sites), which is reproduced below.



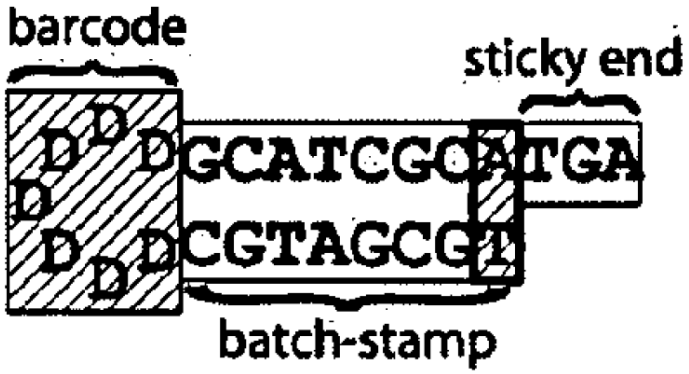
EX1010(Chee), Fig. 1. The amplification sites allow sequencing primers to attach to the nucleic acid to start synthesis of a complementary strand.

81. The Lexow reference shows examples where a tag can contain a variety of different identifiers, including a unique identifier corresponding to the DNA fragment, a molecule identifier corresponding to the class of the molecule, and a sample identifier corresponding to the sample (*e.g.*, patient). EX1018, [0028]-[0029].



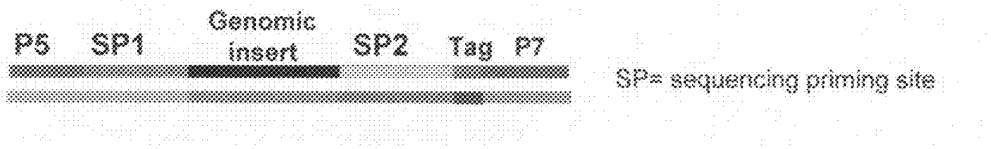
EX1018, Figs. 1 and 2. In the figures above, the 1s and 0s are different oligonucleotide sequences. *Id.*, [0044].

82. McCloskey also discloses tags that include a unique identifier (barcode) corresponding to the molecule and a batch-stamp corresponding to the sample. This is shown in the figure below.



EX1012(McCloskey), Fig. 3.

83. Bignell shows that tags can include sample identifiers. A nucleic acid with a tag and several adaptors is shown below.



EX1013(Bignell), Fig. 5.

84. Prior to May 2009, it was known to use tags that included an identifier corresponding to the sample, and to apply the same tag to every nucleic acid in a sample. This could allow samples to be mixed, while allowing each sequenced nucleic acid to be associated with its sample. During multiplex sequencing, the sample identifier tag allows each fragment to be associated with its sample. EX1018, [0029]; EX1013(Bignell), 1:24-28, 9:53-58; EX1029, [0007], [0014]. It was also known that the identifier could correspond to a class of molecules or a particular gene.

85. It was a known use of tags to apply a different, unique identifier to every nucleic acid in a sample prior to amplification. Use of such unique molecule identifiers could allow each nucleic acid, and any copies of that nucleic acid to be tracked. For example, Chee teaches that "'unique'… identifiers are associated with individual molecules in the starting sample[,] [and] [t]herefore, any amplification products of these initial individual molecules bearing the unique identifier are assumed to be 'identical by descent.'" EX1010(Chee), 5:31-36.

McCloskey teaches that adding a "barcode" to a nucleic acid "can uniquely identify a target nucleic acid" and it "provide[s]… a unique identity to each target nucleic acid in the sample prior to amplification." EX1012(McCloskey), [0021]. Lexow teaches that "individual molecules… can be tracked throughout an analysis procedure by using a molecular tag that is unique to each specific molecule." EX1018, [0005].

86. Not only were the many uses of unique tags in DNA sequencing processes well-known, but the process of making tags with unique sequences had been known for over a decade. In the late 1990s, Sydney Brenner described multiple techniques for creating unique barcodes and attaching them to nucleic acids prior to amplification of a sample. Brenner described using the barcodes to sort the amplified sample and group nucleic acids that originated from the same molecule. EX1019, Abstract; EX1020, 1665-68.

V. ANALYSIS OF THE '730 PATENT

A. Overview of the '730 Patent

87. The '730 patent generally describes a method of correcting for errors or bias in the nucleic acid sequencing workflow. The challenged '730 claims recite the known feature of adding unique tags in the nucleic acid sequencing workflow and utilizing known methods to correct for errors or bias.

88. These concepts were well-known to those of ordinary skill in the art at the time, and these terms are not used in a manner inconsistent with the understanding of a person of ordinary skill.

B. Construction of Terms Used in the '730 Patent Claims

1. "Set of Differentiator Tags" / "Set of Unique Differentiator Tags"

89. Claim 1 recites "s*et of differentiator tags*" and claim 9 recites "*set of unique differentiator tags*." These two terms both refer to a set of prepared nucleic acids where each nucleic acid has a unique sequence. EX1001, 28:53-57.

90. The specification provides several descriptions of differentiator tags. Those descriptions show that the sequence of each differentiator tag is inherently unique such that each target molecule has a different differentiator tag and is uniquely identifiable. For example, the specification describes "associating unique sequence tags (referred to as differentiator tag sequences) with individual target molecules... These tags are useful to distinguish independent target molecules from each other." EX1001, 28:25-32. The specification also states that "a 'differentiator tag sequence'… identifies a unique, independently isolated nucleic acid." *Id.*, 28:53-57.

91. Applicant applied this same definition during prosecution to distinguish the claims from cited prior art that disclosed use of tags. During prosecution, the Applicant argued that the Examiner rejected the claims over prior

art that showed use of sample tags. *E.g.*, EX1003, 1028 ("Ronaghi reports tags to identify the BAC clone from which a sequence was obtained in order to aid sequence assembly of adjacent fragments in the genome. Ronaghi uses a unique pentamer to identify all the clones from a given BAC."); *also id.*, 1176-78. The Applicant also argued that the Examiner rejected the claims over prior art that used probes that had a unique sequence corresponding to particular genes, such that every copy of the gene in the original sample would have the same unique sequence. *E.g.*, EX1003, 1027-28 ("Pourmand is directed to locus detection… There is one tag per locus, and the tags are used to track the locus. The barcode identifies the specific gene… Furthermore, the examples provided by Pourmand confirm that… all candidate probes have the same tag sequence for a given locus or genome and each locus or genome has a different and distinct tag.").

92. To distinguish the art, the Applicant amended the claims and explicitly argued that the claims require every molecule in the sample to have a different, unique tag. EX1003, 1266 ("Thus, in contrast to the present invention, Ronaghi reports batch tagging wherein more than one originating nucleic acid molecule shares the same tag. Ronaghi' s methods of batch tagging are well suited to identifying a library but one of skill in the art would have found no suggestion to modify the tags to be unique to each molecule…"). Applicant's amendment to get

around the sample- and gene- specific tags is below. This claim language eventually became claim 1 of the '730 patent:

> 68. (Currently Amended) A method for correcting for errors or bias introduced during nucleic acid analysis workflow~~identification~~, the method comprising the steps of:
>
> obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
>
> introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;
>
> ~~tagging each of the plurality of target nucleic acid molecules with a unique differentiator tag, thereby generating a plurality of uniquely tagged target nucleic acid molecule~~;
>
> amplifying each of the plurality of ~~uniquely~~ tagged target nucleic acid molecules to generate ampliconsof each uniquely tagged target nucleic acid molecule;
>
> sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
>
> correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count~~; and~~
>
> ~~identifying said target nucleic acid molecule sequences based on the differentiator tag sequence of each of the sequence reads~~.

EX1003, 1258.

93. The Applicant submitted a declaration from first-named inventor Greg Porreca about the scope of these claims. Dr. Porreca stated "[a] distinguishing feature of the claims in my patent application is that each individual nucleic acid molecule is tagged with a unique differentiator tag prior to any amplification or sequencing steps." EX1004, 965. He also stated that "[e]ach differentiator tag has

its own unique sequence so that each individual target nucleic acid molecule is tagged, and therefore associated, with a differentiator tag." *Id.*, 964.

94. On appeal to the Board, the Applicant made it explicit that each differentiator tag was unique and that no two molecules could have the same tag:

> The Examiner's Answer, at page 37, states that the claims do not require a different tag on every single molecule and therefore Appellant's brief and declaration of Dr. Porreca are not arguing limitations of the claim. The Examiner is simply wrong and is not reading either the claim or the Porreca declaration correctly.

EX1004, 1146.

95. I understand that the inventor's and the Applicant's statements about "differentiator tags" inform "the meaning of the claim language by demonstrating how the inventor understood the invention and whether the inventor limited the invention in the course of prosecution, making the claim scope narrower than it would otherwise be." *Phillips v. AWH Corp.*, 415 F.3d 1303, 1317 (Fed. Cir. 2005).

2. "Locus of Origin"

96. Claims 1 and 9 use the terms "*locus of origin*." In the field of genetics, a "locus" refers to a fixed location on a chromosome which contains a gene or other genetic marker. This is consistent with the '730 specification, which describes a genetic loci as a location on the genome that can correspond to

heritable disorders. EX1001, 3:2-20, 3:43-47, 66:3-11; *also id.*, 31:50-42:12 (Table 1).

97. The '730 patent provides examples of genetic loci in Table 1. EX1001, 31:50-42:12. An excerpt from Table 1 is below.

TABLE 1-continued

Target Nucleic Acids

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, *S. cerevisiae*)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAP1; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTD	686	Biotinidase	BTD	609019	NG_008019.1	3p25
HLCS	3141	holocarboxylase synthetase (biotin-(proprionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-P13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2

98. A POSA would have understood that the "locus of origin" of a nucleic acid refers to the location on a chromosome which contains the sequence that encodes for a given nucleic acid molecule.

99. Claim 1 specifies that "*one or more of said loci of origin are associated with more than one differentiator tag*." This means that for at least one locus, the sample includes multiple nucleic acid molecules from that locus, and each nucleic acid molecule from that locus is associated with a different differentiator tag.

100. The Applicant added the locus of origin terms in response to the Examiner's finding that "the claims encompass the same unique tag on every sequence on a sample." EX1003, 1263. The Examiner had found that the claims were unpatentable over prior art that used probes that had a unique sequence corresponding to particular genes, such that every copy of the gene in the original sample would have the same unique sequence. *E.g.*, EX1003, 1027-28 (Applicant distinguishing the claims from Pourmand, a prior art reference, arguing that "Pourmand is directed to locus detection… There is one tag per locus, and the tags are used to track the locus. The barcode identifies the specific gene… Furthermore, the examples provided by Pourmand confirm that… all candidate probes have the same tag sequence for a given locus or genome and each locus or genome has a different and distinct tag.").

101. In response, the Applicant amended the claims to specify that the target nucleic acids come from more than one locus of origin, and that "*one or more of said loci of origin are associated with more than one differentiator tag*."

EX1003, 1258. The Applicant stated that the limitation was added "to make clear that target nucleic acids (loci) may be associated with more than one tag." *Id.*, 1263. According to the Applicant, the loci of origin language clarified that each molecule in the starting sample was associated with a different unique differentiator tag, even if multiple molecules were from the same genetic locus of origin. *Id.*

102. The language of Claim 9 is similar but not identical to Claim 1. Claim 9 specifies that the "*unique differentiator tags are randomly associated with members of said plurality of nucleic acids, such that any given tag is uniquely associated with one locus of origin*." This language excludes the use of tags that are applied to multiple loci, such as sample identifiers. The language requires that each differentiator tag be associated with just one target nucleic acid from one locus of origin.

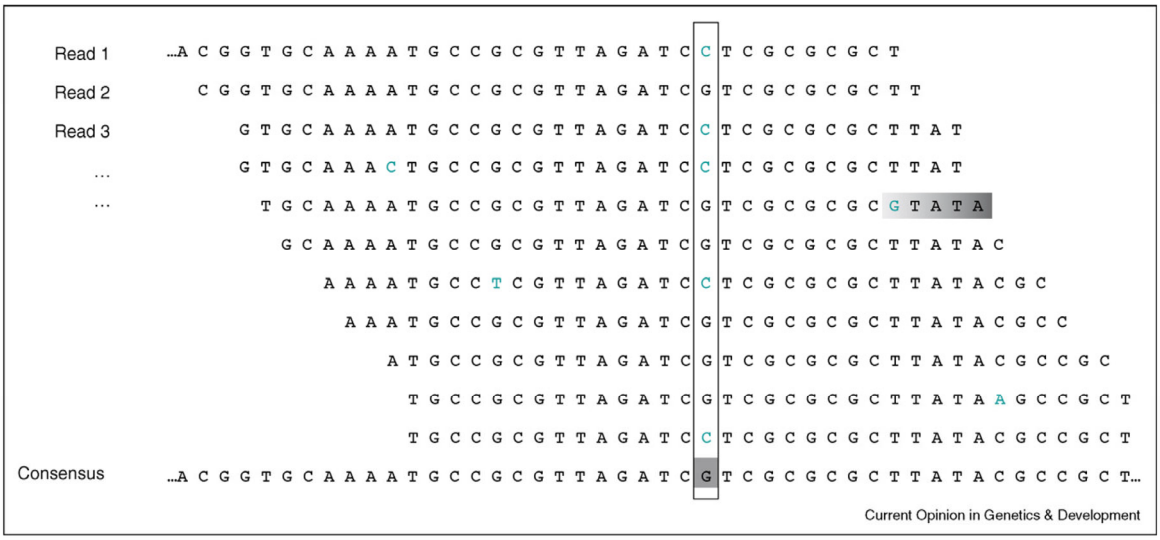
3. "Consensus Base Call"

103. Claim 2 recites "*determining a consensus base call for the target nucleic acid molecule sequences*." A POSA would have been familiar with what a consensus base call is, as that term was used in the art prior to May 2009. Its ordinary meaning in the art referred to using multiple sequencing reads of the same molecule to determine the nucleotide (base) at each position of a sequence.

104. The ’730 patent does not define or describe what a “*consensus base call*” is. Instead, it appears to use the term consistent with its ordinary meaning.

105. When sequencing DNA, it was common to make multiple reads of the same segment of DNA. In that context, the multiple reads can be used to detect and correct errors caused by amplification or sequencing. For some positions in the DNA sequence, there can be differences in the nucleotide (or base) that is identified at that position. Scientists resolve those inconsistencies to come up with a single “consensus” base call for that nucleotide. *See, e.g.*, EX1017(Bentley), 549 (“multiple reads from the same sample can be aligned and used to derive a consensus base-call at each position”); EX1014, 1142. For example, the figure below demonstrates how the alignment of several sequence reads can be used to generate a consensus sequence, showing the consensus base call “G” at the bottom of the vertical rectangle.

Figure 2



EX1017(Bentley), 550.

106. The multiple reads can be combined to reach a single "consensus" sequence for the DNA segment. This is shown in the bottom row of Figure 2 of the Bentley article (reproduced above).

VI. IDENTIFICATION OF THE PRIOR ART

107. ***Exhibit 1010 ("Chee")*** is a copy of U.S. Patent No. 9,085,798 titled "Nucleic Acid Constructs and Methods of Use." It claims the benefit of Provisional Application No. 61/174,442 which was filed on April 30, 2009. I understand that Chee is prior art to the challenged claims since the provisional application provides support for the claims in Chee.

108. ***Exhibit 1012 ("McCloskey")*** is a copy of U.S. Patent Publication No. 2007/0020640 titled "Molecular Encoding of Nucleic Acid Templates for PCR and

Other Forms of Sequence Analysis." It was filed on July 15, 2007. I understand that McCloskey is prior art to the challenged claims.

109. ***Exhibit 1013 ("Bignell")*** is a copy of U.S. Patent No. 8,053,192 titled "Methods for Indexing Samples and Sequencing Multiple Polynucleotide Templates." It was filed on February 4, 2008. I understand that Bignell is prior art to the challenged claims.

110. ***Exhibit 1015 ("Gahan")*** is a copy of a publication titled "Circulating Nucleic Acids in Plasma and Serum." It was published in the Annals of New York Academy of Sciences in 2008. I understand that Gahan is prior art to the challenged claims.

111. ***Exhibit 1017 ("Bentley")*** is a copy of a publication titled "Whole-Genome Re-Sequencing." It was published in Science Direct's Current Opinion in Genetics & Development in October 2006. I understand that Bentley is prior art to the challenged claims.

112. ***Exhibit 1033 ("Chiang")*** is a copy of a publication titled "High-Resolution Mapping of Copy-Number Alterations with Massively Parallel Sequencing." It was published in Nature Methods in January 2009 and was published online November 30, 2008. I understand that Chiang is prior art to the challenged claims.

VII. ANALYSIS OF THE PRIOR ART AND '730 CLAIMS

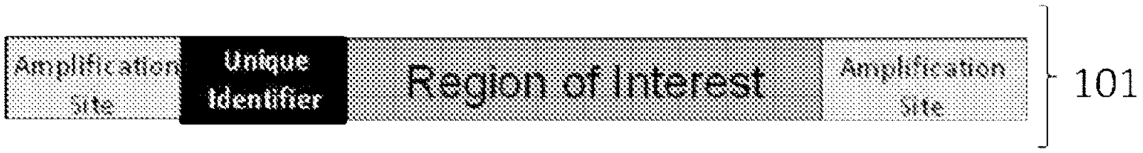
A. Chee Anticipates Claims 1-4 and 9

1. Overview of Chee

113. Chee describes a method of identifying and correcting errors or bias that can arise during the workflow of nucleic acid sequencing. EX1010(Chee), 1:44-58, 2:14-20. For instance, Chee explains that the invention "addresses three primary sources of error encountered in conventional synthetic biology and analysis of nucleic acids in biological samples—that is, errors in nucleic acid synthesis errors introduced in amplification of the original nucleic acid starting materials, and/or errors in sequence identification or determination caused by limitations of analytical methods." *Id.*, 10:21-28; *also id.*, 1:43-47.

114. Chee states that its method can be used in several applications. *See, e.g.*, 5:3-5, 10:21-28, 12:17-22, 8:6-11, 1:11-12. One application its method can be used for is the analysis of synthetic nucleic acids (*i.e.,* in "synthetic biology"). EX1010(Chee), 5:3-5. Another application is the analysis of biological samples, such as DNA. *Id.*, 11:51-54; *also id.*, 5:3-5 ("diagnostic methods"). As Chee explains, its method can be used with the analysis of "naturally-occurring nucleic acids (*e.g.*, DNA or RNA or fragments thereof that are isolated from a sample)." *Id.*, 11:51-54. Thus, Chee's techniques can be used to analyze both biological and synthetic nucleic acid samples.

115. Chee's method involves creating "a set of nucleic acid constructs, wherein the constructs of the set comprise an ROI, a unique identifier, and an amplification site." EX1010(Chee), 2:34-38; *also id.*, 10:41-46, 3:11-16, 4:1-3, 5:27-36. Chee provides the following example of a construct, shown below, where each labeled region is a sequence of nucleic acids. *Id.*, Fig. 1 (below).



Chee shows the construct can include multiple unique identifiers and additional amplification sites. *Id.*, Figs. 1-5. The constructs comprising the ROI, unique identifier, and amplification site are created prior to amplification. *Id.*, 5:27-36, 21:61-64, Fig. 7.

116. One component of Chee's construct is an ROI, which is a "region of interest." EX1010(Chee), 2:12. Chee typically uses the term ROI or region of interest to refer to a specific target nucleic acid (*e.g.*, a DNA fragment) present in an initial sample. *Id.*, 4:1-3. An ROI "can be isolated from a natural… source." *Id.*, 3:51-55.

117. Each of Chee's constructs has a "unique identifier." Chee explains that a "unique identifier" is an "identifier unique to the individual construct." EX1010(Chee), 2:21-23. Chee further explains that "the unique identifier is

associated with a construct, not the ROI." *Id.*, 2:23-25. This means that each construct has a different unique identifier—regardless of the nucleic acid ROI. If the initial sample contains multiple nucleic acid molecules with the same or substantially the same sequence (*e.g.*, because a blood sample contained multiple copies of the person's DNA), each molecule will have a different unique identifier. As a result, Chee's "set of constructs comprises at least two constructs with substantially the same ROI and different unique identifiers." *Id.*, 2:36-38. This would happen if the starting sample has multiple nucleic acids that correspond to the same segment of DNA.

118. Chee shows that each construct has "one or more amplification sites to allow amplification of the construct and/or the ROI." EX1010(Chee), 2:28-30. One example that Chee provides is a construct with two amplification sites "flanking the ROI and the unique identifier," such as the one shown above in paragraph 115. Chee explains that "amplification sites" are regions which are complementary to primers used in amplification reactions, such as PCR. EX1010(Chee), 13:33-37, 13:40-41. Amplification sites are included in the construct so that both the "unique identifier and the ROI" are amplified in an amplification reaction. *Id.*, 2:67-3:3.

119. Following amplification, Chee "[sequences] the complete constructs []to confirm both the ROI and the unique identifier of a particular construct."

EX1010(Chee), 14:3-5. Thus, Chee describes a method in which both the ROI and unique identifier are sequenced.

120. Chee also explains that the practice of its method may employ conventional techniques with sequencing technology "which are within the skill of those who practice in the art." EX1010(Chee), 8:6-11. I understand that Chee incorporates by reference several references which describe these "conventional techniques," including a laboratory manual, titled "*Genetic Variation: A Laboratory Manual*." EX1010(Chee), 8:18-37 (citing "*Genetic Variation: A Laboratory Manual*" (EX1016)).[1] This manual describes several techniques for different aspects of the nucleic acid workflow. For instance, Chapter 9 provides protocols titled "Preparing Genomic DNA from Whole Blood: Small- and Mid-Scale Extraction" and "Preparing DNA from Blood: Large-Scale Extraction." EX1016, 103-06, 120-28. Additionally, Part 3 of the manual is titled "Copy Number and Complex Variation Analysis" and provides several chapters describing techniques for analyzing copy number variation. EX1016, ix; *see generally id.*, 221-53.

[1] I am one of the editors of this manual. The manual was published and publicly available prior to May 2009.

2. Chee was Sufficiently Described in the ’442 Provisional Application

121. I understand that for Chee to be entitled to the priority of its provisional application (EX1011), at least one claim of the patent must have written description support in the provisional. I have reviewed the ’442 provisional and found that at least Claims 1 and 2 in the issued Chee patent were sufficiently described by the ’442 provisional. Below, I identify passages of the provisional that map to each claim element.

a) Chee’s Claim 1

(1) Preamble

122. The preamble of Chee’s claim 1 specifies “*[a] method for determining the sequences of a set of nucleic acid targets*.” There is support in the ’442 provisional for the preamble. For instance, the ’442 provisional states that “complete constructs are sequenced to confirm both the ROI and the unique identifier.” EX1011, [00080]; *see also id.*, [00048], [00061], [00081]. A “complete construct” consists of a ROI (*i.e.*, the nucleic acid target). Determining the sequences of complete constructs therefore refers to “*determining the sequences of a set of nucleic acid targets*.”

(2) tagging each individual nucleic acid target in a set of nucleic acid targets to create a set of tagged constructs, with: (a) at least one unique identifier oligonucleotide that distinguishes the tagged construct from other tagged

constructs in the set, wherein the at least one unique identifier oligonucleotide is not part of the nucleic acid target; and (b) at least one amplification site;

123. There is support in the '442 provisional for this limitation. *See, e.g.*, [0008]-[0009], [00012], [00017], [00022], [00024], [00027], [00055]-[00057], [00060], [00069]-[00070], [00076]. For instance, the '442 provisional states that each construct comprises a "unique identifier," EX1011, [0009], which "is associated with a construct, not the ROI," *id.*, [0008]. The '442 provisional also states that "the different constructs have unique identifiers." *Id.*, [00027]. Further, the '442 provisional teaches that the constructs are added to a sample of isolated nucleic acids containing the ROI. *Id.*, [00069]; *see also id.*, [00017].

124. A person of ordinary skill in the art would have understood that adding the constructs, each having a unique identifier, to a sample of target nucleic acids would serve the function of "*tagging each individual nucleic acid target in a set*." In turn, this would "*create a set of tagged constructs*." *See, e.g.*, [00024]. Additionally, because the constructs have different unique identifiers, a person of ordinary skill in the art would have understood that the unique identifiers could be used to "*distinguish[] the tagged construct from other tagged constructs*."

125. Finally, the '442 provisional states that the "constructs comprise one or more amplification sites." EX1011, [0008]; *see also id.*, [00012], [00022].

(3) amplifying the set of tagged constructs; and

126. There is support in the ’442 provisional for this limitation. A person of ordinary skill in the art would have understood that a construct with a unique identifier and a ROI comprises the “tagged construct.” Thus, a person of ordinary skill in the art would have understood that “amplifying the constructs comprising a unique identifier and a ROI” means “*amplifying the set of tagged constructs*.”

127. The ’442 provisional teaches amplifying constructs with both the unique identifier and the ROI. For instance, the ’442 provisional states that the “constructs allow the unique identifier and the ROI to be selectively amplified.” EX1011, [00012]; *see also id.* [00021]-[00022], [00059]-[00060], [00076]-[00079].

(4) determining the sequences of the amplified set of tagged constructs comprising the unique identifier oligonucleotides and the nucleic acid targets.

128. There is support in the ’442 provisional for this limitation. For instance, the ’442 provisional teaches that the “complete constructs are sequenced to confirm both the ROI and the unique identifier.” EX1011, [00080]; *see also id.*, [00048], [00061], [00081]. It also teaches that the “master set [of constructs] is amplified prior to the sampling and sequencing.” *Id.*, [00059].

b) Claim 2

129. Claim 2 depends on claim 1 and further recites "*wherein the nucleic acid targets are isolated from a biological sample.*" There is support in the '442 provisional for this claim. For instance, the '442 provisional states that the constructs can be "used in quantitative analysis of nucleic acids derived from a biological sample." EX1011, [00069].

3. Chee Describes the Elements of Independent Claims 1 and 9 of the '730 Patent

130. It is my opinion that Chee anticipates claims 1 and 9. Claim 1 (EX1001) is exemplary and reproduced below.

> 1. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:
>
> obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
>
> introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;
>
> amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
>
> sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and

correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

a) Preamble

131. The preamble of claims 1 and 9 specify "[*a*] *method for correcting for errors or bias introduced during nucleic acid analysis workflow*."

132. Chee discloses a method for correcting for errors or bias introduced during a nucleic acid analysis workflow, such as during amplification or when sequencing.

133. As I explained above, Chee explains that its method "addresses three primary sources of error encountered in conventional synthetic biology and analysis of nucleic acids in biological samples—that is, errors in nucleic acid synthesis errors introduced in amplification of the original nucleic acid starting materials, and/or errors in sequence identification or determination caused by limitations of analytical methods." EX1010(Chee), 10:21-28; *see also id.*, 1:43-47, 10:11-14. Chee describes several ways to correct these errors and bias.

134. Chee can correct for bias introduced during amplification processes. For instance, Chee teaches that "the methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier."

EX1010(Chee), 10:41-46; *see also id.*, 10:11-14. Chee recognizes that amplification bias results in the sample having more copies of certain nucleic acids than others, which makes it difficult to determine how many copies of each were in the original sample. Chee explains that use of the unique identifiers solves that problem because copies of the same nucleic acid can be identified. By "tak[ing] into account potential areas of technical bias, including bias in amplification," Chee's methods allow for "the quantification of the molecules [to] be corrected to reflect this manner of bias." EX1010(Chee), 23:53-64; *see also id.*, 9:18-21.

135. Chee's methods can detect and correct errors such as sequencing errors. Chee explains that technical errors can arise whenever complementary strands of DNA are synthesized (*e.g.*, during the sequencing step), and that those errors can be corrected by comparing the sequence reads of multiple molecules with the same unique identifier. EX1010(Chee), 9:29-39. Using these reads, "a consensus sequence can be determined." *Id.*, 14:39-50. As Chee explains, "sequencing error… can be overcome by sequencing multiple instances of the same ROI (i.e. identical by descent)." *Id.*, 10:54-59. The errors are "overcome" because Chee can correct them.

136. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00025], [00060], [00056], [000131].

b) "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin"

137. Claims 1 and 9 specify the above limitation, which Chee teaches.

138. Chee teaches that its method can be used in the analysis of "nucleic acids derived from a biological sample." EX1010(Chee), 12:17-22. Such a biological sample may include "naturally-occurring nucleic acids (*e.g.*, DNA or RNA or fragments thereof that are isolated from a sample)." *Id.*, 11:51-54. For example, nucleic acids are naturally present in blood. Chee's Figure 7 also shows that each sample comprises "*a plurality of target nucleic acid molecules*." *Id.*, 21:61-64, Fig 7 (700).

139. Chee also teaches that the sample comprises target nucleic acids from "*more than one locus of origin*." For instance, Chee explains that its method "is especially useful for very complex sequencing sets or sequencing of nucleic acids having highly related sequences ***in different regions***." EX1010(Chee), 5:24-26; *see id.*, 11:56-60. A person of ordinary skill in the art would have understood that the analysis of nucleic acid molecules "in different regions" refers to nucleic acid molecules at different loci of origin on a genome.

140. Chee describes using its methods for clinical or diagnostic purposes. EX1010(Chee), 11:60-63. When sequencing for a clinical or diagnostic purpose, clinicians would have been looking for sequence variants that were known to

correspond to certain diseases. Examples of such tests from prior to May 2009 include testing gene panels for hereditary diseases such as hearing loss or hypertrophic cardiomyopathy. *See, e.g.*, EX1032, 654 ("16 different genes are implicated in the pathogenesis of hypertrophic cardiomyopathy"). In addition to clinical tests, other applications such as forensics depended on analysis of multiple loci across the genome to try to match an unknown sample with potential matches.

141. A person of ordinary skill in the art would have understood that Chee's method can be used for several purposes which normally involve obtaining and analyzing target nucleic acid molecules from multiple loci.

142. For instance, Chee explains that its method can be used when there are "large number of sequences that need to be determined, such as genomic sequencing." EX1010(Chee), 9:40-42. Chee's method can be used "[w]hen the starting nucleic acid is quite large, e.g., a genome or a complex mixture of nucleic acids." *Id.*, 11:56-60. A person of ordinary skill in the art would have understood that these types of "genomic sequencing" analyses would use samples that included nucleic acid molecules from multiple loci of origin.

143. Chee discloses that its technique for "genomic sequencing" can be used to identify "polymorphisms[,] mutations," or "a rare sequence variant." EX1010(Chee), 9:42-52. In general, Chee's sequencing method could be used for "discovery research, diagnostic use, clinical use or other development techniques."

Id., 11:60-63. A person of ordinary skill in the art would have understood that such diagnostic and clinical use, such as detecting polymorphisms, mutations, and sequence variants, would have involved obtaining and analyzing nucleic acids from multiple loci of origin across the genome.

c) Claim 1: "introducing a set of differentiator tags wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag"

144. Claim 1 specifies the above limitation, which Chee teaches.

145. Chee explains that the unique identifiers and amplification sites are separately synthesized, and it describes several processes for creating them. EX1010(Chee), 11:12-38. The identifiers and amplification sites are added to a sample containing the target nucleic acids ROIs. The identifiers and amplification sites are ligated to the ROIs, forming the constructs. *Id.*, 2:34-38, 2:65-3:9, 3:61-63. As Chee states, "the unique identifier and amplification regions are added to the individual fragments." *Id.*, 4:1-3; *see also id.*, 12:17-22, Fig. 7 (702).

146. The unique identifiers correspond to the "*differentiator tags*" and each nucleic acid molecule in the sample is associated with one of the unique identifiers.

147. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [0007]-[0009], [00011], [00017], [00068].

148. Chee describes several such techniques for creating the unique identifiers it uses. One method is synthesizing the unique identifiers to have specific, "pre-defined" sequences. EX1010(Chee), 11:18-20 ("a diverse but pre-defined set of sequences can be synthesized and attached to the ROIs"). This technique would be labor intensive, but would result in a set of completely unique identifiers. This process might be used when the number of molecules to be tagged is relatively small. Chee also describes a second method where unique identifiers are created such that each is generated by randomly adding bases until a certain length is reached. *Id.*, 11:14-18. This method would be less expensive, and it might be used when a relatively larger number of identifiers is needed, and if oligo synthesis capacity is limited.

149. Chee states that "the term 'unique' as used herein does not necessarily mean absolutely unique but rather having an extremely high probability of being unique." EX1010(Chee), 9:60-63. When the unique identifiers are created by randomly adding bases, there is a chance that the same sequence could be synthesized twice. After creating a large number of identifiers, a subset is selected, and the result is a set of identifiers that has a high probability of being unique. As Chee states, it creates "a large excess of possible identifiers relative to the ROIs in a given set," which makes the chances of having one unique identifier associated with more than one nucleic acid "extremely low." *Id.*, 9:54-60; *see id.*, 11:36-38.

150. Although Chee states that "unique… does not necessarily mean absolutely unique," EX1010(Chee), 9:60-63, Chee discloses both use of identifiers that are absolutely unique and use of identifiers that are only unique with a high probability. A person of ordinary skill in the art would have recognized that Chee discloses both types of identifiers, and either one could be used in Chee's method. The choice of which type of identifier is left the person or team designing the analysis procedure. In other words, a person of ordinary skill in the art would have understood Chee to disclose a method that uses identifiers that are absolutely unique.

151. Chee recognizes that various parameters of its method will be adjusted or changed by the scientists implementing it. For example, Chee notes that people implementing its method can choose the appropriate number of identifiers to achieve their purpose, and explains that in practice, the "numbers of constructs in sets of the invention will depend on the current technology and the ultimate purpose of the set analysis." EX1010(Chee), 2:62-64.

152. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [0007]-[0009], [00011], [00055], [00057], [00064], [00066], [00074].

153. Each nucleic acid molecule in the sample is associated with a different unique identifier. Chee states that each individual construct has its own unique identifier allowing it to be "uniquely identifiable." EX1010(Chee), 2:10, 2:21-28.

Each construct also has one ROI, which is a nucleic acid molecule from the starting sample. The result is that each nucleic acid molecule in a starting sample is part of its own construct and with its own unique identifier. As Chee explains:

> It is a feature of the invention that the ***different constructs have unique identifiers***, including constructs that may have the same ROI, so the specific, correct ROIs can be ***distinctly identified*** and distinguished from the other constructs in the set. In essence, ***the 'unique'*** (i.e. unique for a construct comprising a particular ROI) ***identifiers are associated with individual molecules in the starting sample***. Therefore, any amplification products of these initial individual molecules bearing the unique identifier are assumed to be 'identical by descent'.

EX1010(Chee), 5:27-36.

154. In the passage above, Chee states that "the 'unique' (*i.e.*, unique for a construct comprising a particular ROI) identifiers are associated with individual molecules in the starting sample." EX1010(Chee), 5:31-33. In other words, no two nucleic acids in the starting sample will be associated with the same unique identifier.

155. Chee explains that the initial sample may include multiple nucleic acids from the same region of a genome. Chee states that its method can create multiple constructs that have "substantially the same ROI" but different unique identifiers. EX1010(Chee), 2:34-38 ("two constructs with substantially the same

ROI and different unique identifiers"), 2:38-43 (describing 5 or 10 such constructs). By "substantially the same ROI," Chee means nucleic acids from the same locus of origin on the genome (*e.g.*, because the sample naturally included multiple copies of the same DNA). These nucleic acids would have the same or overlapping sequences. Because each nucleic acid molecule is part of its own construct, each one is associated with a different unique identifier. Any copies made of it can be identified. This allows an original molecule, and any copies made of it, to be identified and traced throughout the sequencing process. EX1010(Chee), 10:41-46, 24:1-4.

156. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [0007]-[0009], [00011], [00027], [00029], [00055]-[00056], [00060]-[00061].

d) Claim 9: "introducing a set of unique differentiator tags wherein members of said set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin"

157. Claim 9 specifies the above limitation, which Chee teaches. During prosecution, Applicant made clear that "*differentiator tags*" must be unique. Therefore, including the word "*unique*" in this claim therefore does not add meaning to the term "*differentiator tag*." Chee's unique identifiers are "*unique*

differentiator tags" for the same reason they are "*differentiator tags*" as discussed above with respect to claim 1.

158. This limitation is similar to claim 1, but further requires that the "*set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules*." The differentiator tags in Chee, referred to as "unique identifiers," do not have a region that specifically targets a ROI, and each unique identifier is able to attach to any nucleic acid in the sample. *See* EX1010(Chee), 4:1-3, 11:12-18. Chee also discusses how to create random identifiers, and statistically select identifiers so that the probability that any two constructs having the same identifier is small. *Id.*, 11:26-38. Therefore, a person of ordinary skill in the art would have understood that the tags randomly associate with the target nucleic acid molecules.

159. This limitation also further requires that "*any given tag is uniquely associated with said locus of origin*." Each of Chee's unique identifiers is associated with one ROI, and therefore, each is associated with just one locus of origin. Chee explains that "[i]t is a feature of the invention that the different constructs have unique identifiers, including constructs that may have the same ROI." EX1010(Chee), 5:27-31; *also id.*, 5:31-36, Fig. 1. This means that every ROI will be uniquely associated with a unique identifier.

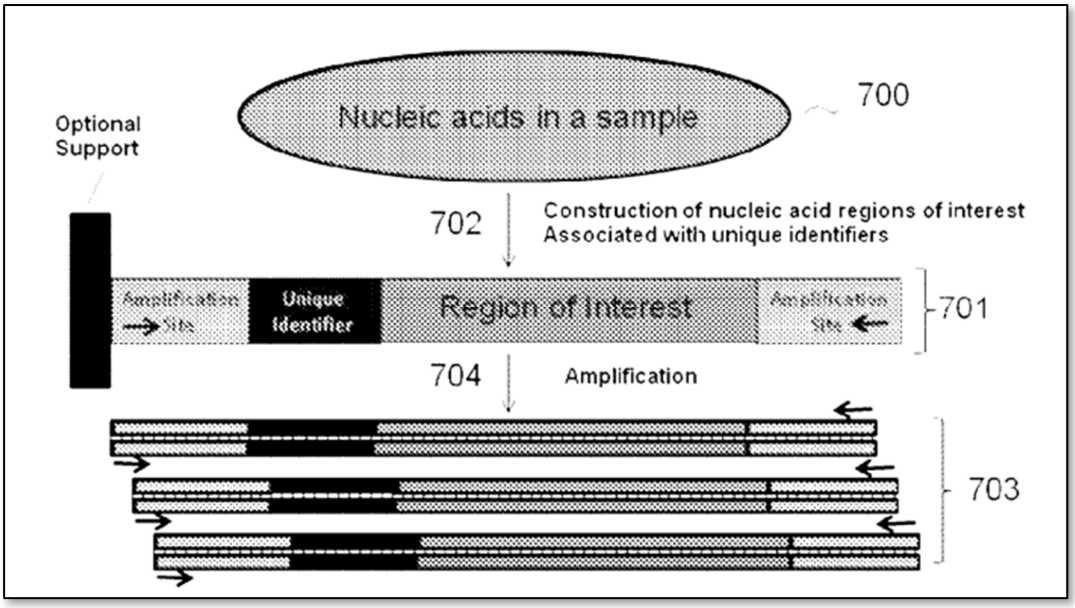
160. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00017], [00027].

e) "amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons"

161. Claims 1 and 9 specify the above limitation, which Chee teaches.

162. Chee teaches that the tagged target nucleic acid molecules are amplified to generate an amplified sample. The amplified nucleic acids can be called amplicons. For instance, Chee explains that "[t]hese constructs allow the unique identifier and the ROI to be selectively amplified using conventional, two-primer amplification techniques such as the polymerase chain reaction (PCR)." EX1010(Chee), 2:65-3:3; *see id.*, 13:20-61, 13:33-37 ("each unique molecule now has a small number of identical copies each containing an ROI and the unique identifier"), 21:64-67. Chee explains that its constructs can "compris[e] a single amplification site positioned to allow amplification of the ROI and the unique identifier." *Id.*, 3:10-17.

163. Figure 7, below, depicts how tagged nucleic acid molecules (*i.e.*, the ROI and the unique identifier, marked as 701) are amplified (marked as 704) to generate amplicons (marked as 703). *See* EX1010(Chee), Fig. 7 (below).



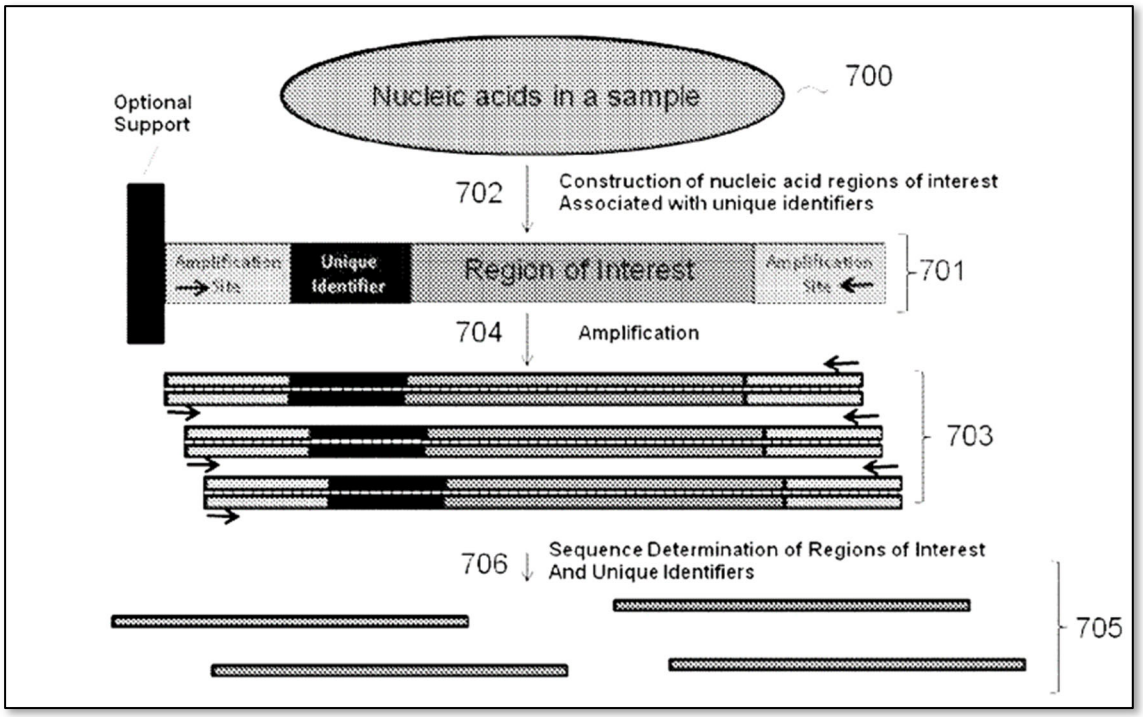
164. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00012]-[00013], [00076]-[00078].

f) "sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence"

165. Claims 1 and 9 specify the above limitation, which Chee teaches.

166. Chee teaches that the "*amplicons obtained in said amplifying step*" are sequenced. Chee states that "complete constructs are sequenced to confirm both the ROI and the unique identifier of a particular construct." EX1010(Chee), 14:3-10; *also id.*, 14:42-59. Chee describes analyzing the sequence of both the ROI and unique identifier. *Id.*, 9:25-39, 14:42-50.

167. Chee depicts this process in Figure 7, and states that "[t]he sequence is then determined 706 for the pool of amplified constructs 705." EX1010(Chee), 21:67-22:1. Figure 7 shows that the amplicons obtained in the amplification step are sequenced and the sequence reads comprise the region of interest sequence and the unique identifier sequence. The sequencing step in Figure 7 is titled "Sequencing Determination of Regions of Interest And Unique Identifiers." *See* EX1010(Chee), Fig. 7 (below).



168. Further, Chee teaches that "*each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence*." In Chee's method, "complete constructs are sequenced to confirm both the ROI and the unique identifier of a particular construct." EX1010(Chee), 14:3-5. A person

of ordinary skill in the art would have understood that sequencing the ROI (*i.e.*, *target nucleic acid molecule*) and the unique identifier (*i.e.*, *differentiator tag*) would generate sequence reads comprising the sequences of both the ROI and the unique identifier.

169. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00055], [00080], [00083].

g) Claim 1: "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count"

170. Claim 1 specifies the above limitation, which Chee teaches.

171. Chee's method can detect and correct for errors and bias caused by the amplification and sequencing processes. EX1010(Chee), 10:41-57, 23:51-59.

172. Chee describes how its methods can identify and correct bias in several places. Chee recognizes that "technical bias" can be introduced when amplifying the original sample, which can "skew" the final sample. EX1010(Chee), 9:18-21; *see id.*, 10:36-46. As I noted above, this can result in the final sample including more copies of a particular construct as compared to other constructs.

173. Chee explains that the unique identifiers allow it to determine how many of each molecule was present in the initial sample. EX1010(Chee), 10:41-46. Chee uses the unique identifiers to determine that two sequenced constructs

are "identical by descent," by which Chee means they are copies of (or descended from) the same construct from the original sample. *Id.*, 5:27-36, 14:42-50. This use of unique identifiers allows Chee to "correct" the "quantification of the molecules" to fix the bias. *Id.*, 23:59-64. By correcting the quantification, Chee means that where it reads multiple sequences with the same unique identifier, it only counts those as one instance of the ROI (not multiple instances). This "collapses" multiple reads into one count.

174. The process is described in the following two passages from Chee:

> In addition, the methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus ***the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier***.

EX1010(Chee), 10:41-46.

> Improved Quantification of ROIs
>
> Methods to determine the relative amounts of nucleic acids from a sample often require amplification methods in order to have enough material to carry out the desired testing. Conventional methods do not take into account potential areas of technical bias, including bias in amplification of certain sequences. By creating constructs comprising the nucleic acids directly isolated from a sample, bias due to subsequent manipulation of the molecules can be identified by determining the amounts of constructs with unique identifiers. Thus, ***if***

> ***a nucleic acid is either over-represented or under-represented due to amplification bias***, the number of molecules comprising unique identifiers will also be over-represented or under-represented, and ***the quantification of the molecules can be corrected to reflect this manner of bias***.

EX1010(Chee), 23:50-64; *see also id.*, 9:5-9, 24;1-8.

175. The '730 patent shows that its method performs the "*collapsing*" step by counting the number of unique tags associated with each sequence (and ignoring duplicates) to determine the number of times a sequence appeared in the original sample. EX1001, 30:48-51, 72:63-73:5. That same process is shown by Chee.

176. Chee describes how its methods can identify and correct errors in several places. Chee observes that errors can be introduced when synthesizing copies of nucleic acids, and that such errors can be overcome (corrected):

> The second class of errors addressed by the constructs and methods of the invention includes errors that are introduced from the enzymatic amplification used to increase the amount of shorter nucleic acid fragments created through chemical synthesis. In addition, error can be introduced as a result of sequencing error; this can be overcome by sequencing multiple instances of the same ROI (i.e. identical by descent).

EX1010(Chee), 10:50-57.

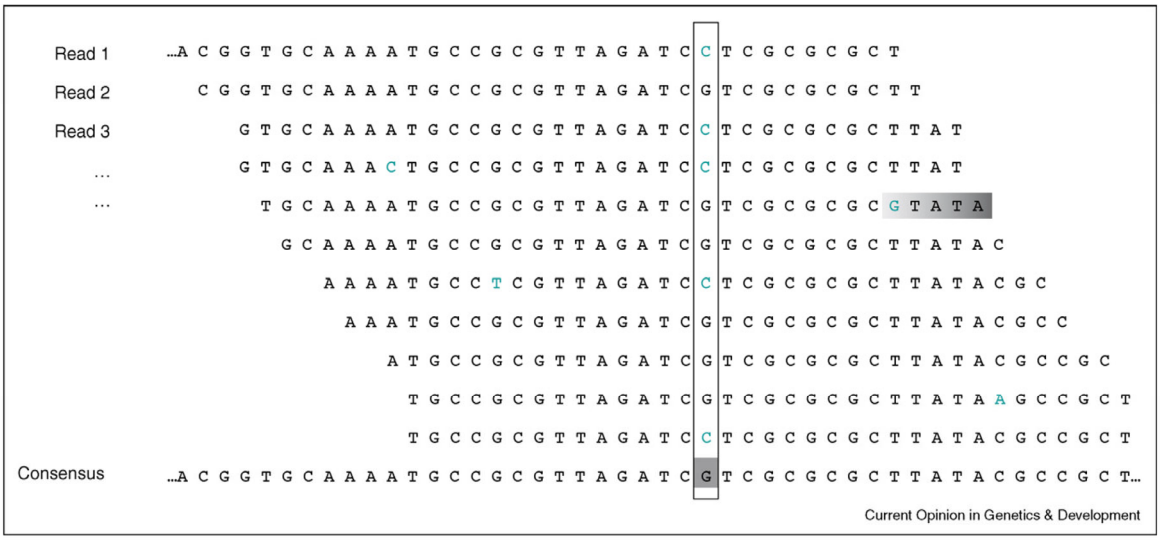
177. Chee explains that those errors can be corrected by sequencing multiple copies of the same original molecule and comparing them. EX1010(Chee), 9:34-39, 10:54-57, 14:39-50. Chee can identify multiple copies of the same original molecule using the unique identifier. *Id.*, 10:54-57, 14:39-50. Chee describes using the well-known concept of a consensus sequence or a consensus base call to correct any such errors. As Chee explains:

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

EX1010(Chee), 14:42-50.

178. It was well-known to those in the art that deriving a consensus sequence involved considering sequencing reads from multiple overlapping copies of a DNA sequence to come up with a single sequence for the DNA segment. Sometimes errors can arise from amplification or sequencing. For a particular base, if one sequencing read is different than the other reads for that base, the errant read is likely an error. That process is described in a 2006 article by David Bentley, and depicted in the figure below:

Figure 2



EX1017(Bentley), 550.

179. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00027], [00055], [00058]-[00060], [00083].

h) Claim 9: "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule"

180. Claim 9 specifies the above limitation, which Chee teaches. Chee describes a method of "*correcting for error or bias introduced during said workflow*," as I discussed above with respect to claim 1.

181. This limitation is similar to claim 1, but the limitation differs slightly in that it requires "*recognizing target/differentiator tags having the same sequence as being derived from the same input molecule*." Chee recognizes constructs with the same unique identifier as being "identical by descent." For instance, Chee

explains that "any amplification products of these initial individual molecules bearing the unique identifier are assumed to be 'identical by descent.'" EX1010(Chee), 5:27-36. A person of ordinary skill in the art would have understood that the phrase "identical by descent" referred to amplified constructs which originated from the same initial construct or as "*being derived from the same input molecule*." Thus, the amplified constructs share the same unique identifier and should theoretically share the same ROI sequence. A person of ordinary skill in the art would have understood that differences in the ROI sequences among molecules that are "identical by descent" indicate error or bias. Chee also explains that "sequence differences between molecules with the same ROI can be identified as technical errors." EX1010(Chee), 9:29-34.

182. Chee teaches that the sequences of constructs that are "identical by descent" can be compared to correct for error and bias by creating, for instance, a consensus sequence. Chee explains that a "consensus sequence [is] derived from multiple clonal copies of the construct, which are identified as identical by descent via the shared unique identifier." EX1010(Chee), 14:46-50.

183. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00027], [00059]-[00061], [00083].

4. Claim 2

184. Claim 2 depends on claim 1 and further recites "*wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences*."

185. Chee teaches this limitation. As I explained above with respect to claim 9, Chee teaches that sequencing errors "can be overcome by sequencing multiple instances of the same ROI." EX1010(Chee), 10:54-57. Chee's method for correcting errors includes generating a "consensus sequence derived from multiple clonal copies of the construct, which are identified as identical by descent via the shared unique identifier." EX1010(Chee), 14:46-50. The concept of "identical by descent" means that the sequence reads comprise the same unique identifier (*i.e.*, "*differentiator tag*"). Therefore, a person of ordinary skill in the art would have understood that Chee describes "*collapsing all sequence reads comprising a same differentiator tag into a single read*" to determine a consensus base call at each position in the "*target nucleic acid molecule sequences*."

186. Chee's '442 provisional also describes this subject matter. *E.g.*, EX1011, [00061], [00083].

5. Claim 3

187. Claim 3 depends on claim 2 and further recites "*further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences*."

188. Chee teaches this limitation. As I explained above with respect to claim 2, Chee describes "*determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences*" since sequence reads with the same unique identifier (*i.e.*, "*differentiator tag*") are compared to identify and correct the error. EX1010(Chee), 14:46-50; *also id.*, 9:29-34, 14:57-59.

189. Chee teaches that amplification can introduce errors and bias. *See, e.g.*, EX1010(Chee), 9:25-34 (discussing the "limitations of the amplification" technology), 10:21-28 (addressing "errors introduced in amplification"), 10:36-46 ("technical bias"). Chee explains that "the methods of the invention can identify technical bias in, e.g., amplification, as the unique identifier will be amplified with the ROI of each construct and thus the number of starting molecules can be inferred by association of each initial ROI present in a sample with its unique identifier." EX1010(Chee), 10:42-46. This means that, after amplification, the number of different unique identifiers associated with each ROI can be counted to provide information on the number of each ROI that was present in the initial

sample. A person of ordinary skill in the art would have understood that this method corrects for amplification error based on "*a number of differentiator tag sequences*."

190. Similarly, Chee explains that "if a nucleic acid is either over-represented or under-represented due to amplification bias, the number of molecules comprising unique identifiers will also be over-represented or under-represented, and the quantification of the molecules can be corrected to reflect this manner of bias." EX1010(Chee), 23:59-64; *also id.*, 23:65-16 (discussing amplification and error rates). A person of ordinary skill in the art would have understood that this is another method that identifies and corrects for amplification error based on "*a number of differentiator tag sequences*."

191. Chee's '442 provisional also discloses this subject matter. *E.g.*, EX1011, [00058], [00060], [00083].

6. Claim 4

192. Claim 4 depends on claim 1 and further recites "*said biological sample is selected from the group consisting of blood and biopsy tissue*."

193. Chee teaches this limitation. As I explained above, Chee incorporates by reference several references which describe these "conventional techniques," including a laboratory manual, titled "*Genetic Variation: A Laboratory Manual*." EX1010(Chee), 8:18-37 (citing "*Genetic Variation: A Laboratory Manual*"

(EX1016)).[2] This manual describes several techniques for different aspects of the nucleic acid workflow, including the analysis of blood samples and fixed tissue samples. For instance, Chapter 9 provides protocols titled "Preparing Genomic DNA from Whole Blood: Small- and Mid-Scale Extraction" and "Preparing DNA from Blood: Large-Scale Extraction." EX1016, 103-06, 120-28. Chapter 9 also provides a protocol titled "Preparing DNA from Fixed Tissue: Extraction and Whole-Genome Amplification." EX1016, 110-113.

194. A POSA would have understood the conventional techniques described in the *Genetic Variation Manual* are the type of technique that Chee would have used to acquire "nucleic acids in biological samples." EX1010(Chee), 10:24.

195. The '442 provisional also incorporates by reference the *Genetic Variation Manual*. *See* EX1011, [00050].

B. Chee Renders Obvious Claims 1-4, 7, and 9

1. Claims 1 and 9

196. As I explained above in section VII.A.3, Chee discloses every element of independent claims 1 and 9. I understand that Patent Owner may argue that Chee's unique identifiers are different than the claimed "differentiator tags" based

[2] I am one of the editors of this manual.

on Chee's statements that its unique identifiers "need not be absolutely unique" or that unique "does not necessarily mean absolutely unique but rather having an extremely high probability of being unique." EX1010(Chee), 9:60-63, 11:22-25. Although it is my opinion that Chee discloses absolutely unique identifiers, if it did not, a person of ordinary skill in the art would have found it obvious to configure Chee so that they were absolutely unique.

197. Even if Chee's unique identifiers were not unique, Chee at least suggests that its identifiers could be absolutely unique.

198. Prior to May 2009, it was within the skill of those in the art to select appropriate parameters to be used to achieve the purpose of the intended analysis, and Chee explicitly recognizes that. EX1010(Chee), 2:62-64, 8:17-18, 8:67-9:3, 13:5-13, 19:51-56. Chee states that unique identifiers need only have a high probability of being unique because the method will work but will have some small amount of error. The errors would arise because Chee uses the unique identifiers to group the constructs and determine which ones are copies of the same molecule from the starting sample. EX1010(Chee), 10:41-57, 14:42-52. That grouping allows Chee to correct for bias and to correct sequencing errors. If two nucleic acid molecules in the starting sample got the same unique identifier, Chee would be unable to accurately determine if skewing occurred and might incorrectly determine a consensus sequence. A person of ordinary skill would have

recognized that configuring Chee to use identifiers that are absolutely unique would avoid any such errors when grouping constructs. A POSA would have been motivated to use, or try using, identifiers that are absolutely unique to ensure that no two nucleic acid ROIs in a starting sample were associated with the same identifier.

199. A person of ordinary skill in the art would have had a reasonable expectation of success in implementing absolutely unique identifiers in Chee's method as using absolutely unique identifiers would not alter Chee's method. It would only eliminate the chance that two "unique identifiers" would be the same, which Chee recognizes is already extremely low. Techniques for synthesizing unique identifiers were known and within the skill of those in the art. See EX1010(Chee), 11:12-38; EX1019(Brenner), 3:36-65, 7:18-35, 10:41-49.

200. A person of ordinary skill in the art would have been motivated to use absolutely unique identifiers if the number of nucleic acids in a sample was relatively small, or if accuracy was extremely important. Chee teaches two methods of creating unique identifiers. A person of ordinary skill in the art would have been obvious to at least try using Chee's pre-defined method to generate absolutely unique identifiers.

201. Therefore, it would have been obvious to a person of ordinary skill in the art to configure the "unique identifiers" in Chee to be absolutely unique.

2. Claim 2

202. Claim 2 depends on claim 1 and further recites "*wherein said correcting step further comprises collapsing all sequence reads comprising a same differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences*."

203. Chee teaches this limitation. Therefore, Chee renders obvious claim 2 for the same reasons it anticipates claim 2, which I explained above. *See* § VII.A.4.

3. Claim 3

204. Claim 3 depends on claim 2 and further recites "*further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences*."

205. Chee teaches this limitation. Therefore, Chee renders obvious claim 3 for the same reasons it anticipates claim 3, which I explained above. *See* § VII.A.5.

4. Claim 4

206. Claim 4 depends on claim 1 and further recites "*said biological sample is selected from the group consisting of blood and biopsy tissue*."

207. Chee teaches this limitation. As I explained above, Chee incorporates by reference several references which describe these "conventional techniques," including a laboratory manual, titled "*Genetic Variation: A Laboratory Manual*."

EX1010(Chee), 8:18-37 (citing "*Genetic Variation: A Laboratory Manual*" (EX1016)).[3] This manual describes several techniques for different aspects of the nucleic acid workflow, including the analysis of blood samples and fixed tissue samples. For instance, Chapter 9 provides protocols titled "Preparing Genomic DNA from Whole Blood: Small- and Mid-Scale Extraction" and "Preparing DNA from Blood: Large-Scale Extraction." EX1016, 103-06, 120-28. Chapter 9 also provides a protocol titled "Preparing DNA from Fixed Tissue: Extraction and Whole-Genome Amplification." EX1016, 110-113. It therefore would have been obvious to a person of ordinary skill in the art to use these conventional, well-known techniques for extracting DNA from blood and tissue samples with Chee's methods.

5. Claim 7

208. Claim 7 depends on claim 1 and further recites "*further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained*."

209. I reviewed the '730 patent to analyze its disclosure of "copy number." The '730 patent seems to indirectly refer to "copy number" when it explains that "aspects of the invention relate to methods for detecting nucleic acid deletions or

[3] I am one of the editors of this manual.

insertions in regions containing nucleic acid sequence repeats." EX1001, 55:49-51. The '730 patent explains that this aspect of the invention may be useful in analyzing the "[i]nstability in the length of genomic regions that contain high numbers of repeat sequences," which "has been associated with a number of hereditary and non hereditary diseases and conditions." *Id.*, 55:56-59. One example it provides is cancer. The '730 patent explains that cancer "has been associated with microsatellite instability (MSI) involving an increase or decrease in the genomic ***copy number*** of nucleic acid repeats at one or more microsatellite loci (e.g., BAT-25 and/or BAT-26)." EX1001, 56:54-58. This is the only time the '730 patent mentions "copy number."

210. The '730 patent provides no information for how to determine the "*copy number of a genomic region*." It only provides that there "are currently many sequencing-based assays for determining the number of nucleic acid repeats at a particular locus and identifying the presence of nucleic acid insertions or deletions," but that "such techniques are not useful in a high throughput multiplex analysis." EX1001, 56:58-63. Thus, a person of skill in the art would have understood that it should rely on knowledge that was known in the art to determine "*the copy number of a genomic region*."

211. As I explained above, the concept of "copy number variation" ("CNV") refers to an abnormal number of copies of a segment of DNA. An

example of CNV that was known by May 2009 included Trisomy 21 (*i.e.*, Down syndrome). It was also known by May 2009 that CNV is commonly associated with cancer. Thus, a person of ordinary skill in the art would have understood that "*determining the copy number of a genomic region*" could be for diagnostic or clinical purposes.

212. Chee's methods could be used for "discovery research, diagnostic use, clinical use or other development techniques." *Id.*, 11:60-63; *see also id.*, 9:42-52 (discussing identification of polymorphisms, mutations, and rare sequence variants).

213. As I explained above, Chee incorporates by reference several references which describe these "conventional techniques," including a laboratory manual, titled "*Genetic Variation: A Laboratory Manual*." EX1010(Chee), 8:18-37 (citing "*Genetic Variation: A Laboratory Manual*" (EX1016)).[4] This manual states that "[t]he ability to discover and measure changes in chromosome copy number is becoming critical to our understanding of the differences between normal and disease states." EX1016, xvi. This is consistent with how "copy number" determination was recognized by the field at that time. *See, e.g.*,

[4] I am one of the editors of this manual.

EX1014, 1143 (explaining that an "[i]mportant application[]" of next-generation sequencing was identifying "copy number variation").

214. Chee also explains its invention is a method for detecting and correcting bias. A person of ordinary skill in the art would have been familiar with these conventional techniques for analyzing copy number and would have known that copy number analysis can be impacted by bias, such as bias introduced during amplification and sequencing. *See, e.g.*, EX1033(Chiang), 101 (explaining that copy number analysis "may reflect inherent biases in the sample preparation or sequencing procedures" and providing one approach to mitigate these biases). A person of ordinary skill in the art would therefore have been motivated to use Chee's methods for "*determining the copy number of a genomic region*" to mitigate bias and for clinical or diagnostic purposes. It would have been within the skill of a person of ordinary skill in the art to determine copy number using Chee's methods. Chee describes techniques for accurately quantifying the number of molecules that appear in an original sample. EX1010(Chee), 23:50-64. A skilled person would have recognized that technique allowed for accurately determining the number of times a particular sequence appeared in the sample.

215. Further, a person of ordinary skill in the art would have had a reasonable expectation of success in "*determining the copy number of a genomic region*" using Chee's methods. The Genetic Variation manual that Chee

incorporates by reference also provides information on how to analyze and determine copy number. Part 3 of the manual is titled "Copy Number and Complex Variation Analysis" and provides several chapters describing techniques for analyzing copy number variation. EX1016, ix; *see generally id.*, 221-53.

C. Chee and Gahan Render Obvious Claims 5-6 and 8

1. '730 Patent Claims 5-6 and 8

216. Claims 5, 6, and 8 specify that the sample includes nucleic acids from bacteria or a virus, from a tumor, or from a fetus or placenta. Those claims are reproduced below.

> 5. The method of claim 1, wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample.
>
> 6. The method of claim 1, wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules.
>
> 8. The method of claim 1, wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules.

217. The additional limitations of these claims are not novel, as they all were known sources of nucleic acids prior to May 2009. The '730 patent states that its methods can be used when analyzing DNA from the listed sources, but

the ’730 patent does not describe these elements as having any impact on its methods or being of any significance to .

2. Overview of Gahan

218. Gahan is a publication containing a report on the fifth international conference on circulating nucleic acids in plasma and serum, during which recent developments in the analysis of circulating nucleic acids were discussed. EX1015(Gahan), 1. Gahan provides an overview of circulating nucleic acids and describes how circulating nucleic acids may be isolated and used for diagnostic purposes. Gahan states that DNA circulates in and can be isolated from the whole blood, serum, plasma. EX1015(Gahan), 1.

219. Gahan references “[m]any studies [which] have shown the successful application of CNAPS [(circulating nucleic acids in plasma and serum)] in the diagnosis and monitoring of cancer and in prenatal diagnosis, as well as in other diseases.” EX1015(Gahan), 3. Gahan also states that “[t]he ability to isolate, quantify, and analyze these molecules has led to the identification of specific nucleic acid fragments related to particular disorders such as diabetes, cancer, myocardial infarction, and stroke, thereby permitting their early diagnosis.” EX1015(Gahan), 1.

220. One diagnostic purpose is in fetal medicine. Gahan explains that “[t]he analysis of maternal plasma for DNA has been exploited in the early

identification of a range of maternal complications such as preeclampsia and intrauterine growth retardation, as well as in fetal disorders including Rhesus incompatibility, and β-thalassemia." EX1015(Gahan), 3 (internal citations omitted); *also id.*, 2 ("fetal DNA"). Additionally, Gahan reports that "[s]everal studies have shown that the sex of the fetus could be determined using fetal DNA in maternal plasma, many of which have used the presence or absence of Y chromosome markers." *Id.*, 3-4.

221. Gahan also teaches that analyzing circulating nucleic acids can provide diagnostic information in oncology and other diseases. Gahan explains that "the identification of certain fragments of circulating DNA can be used not only as an early diagnostic marker (e.g., in breast, lung, head and neck, and gastrointestinal tumors), but also as a means of following changes during therapy (as in the case of melanoma)." EX1015(Gahan), 4. Gahan also states that "[c]irculating DNA and RNA have been useful as markers of nononcogenic disorders such as trauma, stroke, autoimmune disease, and diabetes mellitus." *Id.*

3. A Person of Ordinary Skill in the Art Would Have Practiced Chee to Analyze Nucleic Acids According to Conventional Techniques

222. A person of ordinary skill in the art would have understood that Chee's method could be used to analyze nucleic acid molecules using conventional techniques. As I explained above, Chee incorporates by reference several

references which describe these "conventional techniques," including a laboratory manual, titled "*Genetic Variation: A Laboratory Manual*." EX1010(Chee), 8:18-37 (citing "*Genetic Variation: A Laboratory Manual*" (EX1016)).[5] This manual describes several techniques for different aspects of the nucleic acid workflow, including the analysis of blood samples and fixed tissue samples. For instance, Chapter 9 provides protocols titled "Preparing Genomic DNA from Whole Blood: Small- and Mid-Scale Extraction" and "Preparing DNA from Blood: Large-Scale Extraction." EX1016, 103-06, 120-28. As explained for claim 4 above, a POSA would have found it obvious to obtain Chee's biological samples from a patient's blood, consistent with the *Genetic Variation Manual*'s teaching of conventional methods for doing so.

223. A POSA would have known that different types of nucleic acids circulate in blood, and that these types of nucleic acids can be specifically studied. That conventional knowledge is reflected in Gahan, which discusses various types of DNA and RNA in blood or other plasmas.

224. A person of ordinary skill in the art in May 2009 would have known that it was useful to sequence DNA circulating in blood for diagnostic purposes. For example, Gahan states such analysis allows for the identification of any nucleic

[5] I am one of the editors of this manual.

acids "related to particular disorders such as diabetes, cancer, myocardial infarction, and stroke, thereby permitting their early diagnosis." EX1015(Gahan), 1. Such analysis of DNA from patient blood samples was routine. It was also well-known that circulating nucleic acids from blood samples could be analyzed for several diagnostic purposes, including for fetal medicine as well as the diagnosis of cancer and other diseases. A POSA would have been motivated to isolate the different types of nucleic acids circulating in blood and analyze them using Chee's method. Such analyses would have included the diagnostic purposes described in Gahan.

4. Claim 5

225. Claim 5 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample*."

226. Chee and Gahan teach this limitation. A person of ordinary skill in the art would have known that bacteria and viruses can be found in human biological samples and that bacteria and viruses have nucleic acid molecules, including DNA and RNA. A person of ordinary skill in the art would have known that bacterial or viral nucleic acids from a human biological sample could be analyzed for diagnostic or clinical purposes. *See, e.g.*, EX1032, 649 ("arenavirus

in transplantation patients"); *id.* ("microflora present in the human oral cavity"); *id.*, 654 (human immunodeficiency virus ("HIV")).

227. A person of ordinary skill would have understood that a patient's blood can include nucleic acids from viruses. Gahan states that viral nucleic acids can be extracted from human biological samples and analyzed using a "QIAamp Ultrasens Virus Kit." EX1015(Gahan), 3. A person of ordinary skill in the art would have known that such a kit can be used to extra viral RNA from blood and that such RNA could be analyzed to identify certain diseases (*e.g.*, HIV). A person of ordinary skill in the art would have been motivated to perform conventional analyses using Chee's methods, such as sequencing nucleic acids from viruses that are circulating in blood.

5. Claim 6

228. Claim 6 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules*."

229. Chee and Gahan teach this limitation. As I explained above, Chee teaches analyzing nucleic acids obtained from a blood sample. A person of ordinary skill in the art would have understood that a blood sample contains circulating nucleic acid molecules. A person of ordinary skill in the art would have also known that tumor cells shed DNA into the bloodstream. Gahan, for instance, explains that "[c]ells releasing these forms of DNAs could include tissue-based

cells related to . . . tumors." EX1015(Gahan), 2; *also id.* (referring also to "cancer-related DNA" from "cancer cells" or "tumors").

230. A person of ordinary skill in the art would have understood that, if a tumor is present, the nucleic acid molecules in blood samples can comprise circulating tumor nucleic acid molecules. *See, e.g.*, EX1032, 654 ("detection of mutations in tumor cells . . . in the circulation"). Gahan reports that analyzing circulating nucleic acids can identify breast cancer and lung cancer tumors, amongst other types of cancer. EX1015(Gahan), 4. Thus, a person of ordinary skill in the art would have been motivated to use Chee's methods to screen for DNA from tumors.

6. Claim 8

231. Claim 8 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules*."

232. Chee and Gahan teach this limitation. As I explained above, Gahan teaches that fetal DNA from maternal blood, specifically blood plasma, can be analyzed. For instance, Gahan reports that "[s]everal studies have shown that the sex of the fetus could be determined using fetal DNA in maternal plasma, many of which have used the presence or absence of Y chromosome markers." EX1015(Gahan), 3-4; *also id.*, 2. This means that a maternal blood sample

comprises circulating fetal cells which can be analyzed, such as to determine the sex of the fetus. Thus, a person of ordinary skill in the art would have been motivated to use Chee's methods to screen fetal DNA.

D. Bignell and McCloskey Render Obvious Claims 1, 4-6 and 9

1. Overview of Bignell

233. Bignell describes methods for tagging nucleic acids in samples and sequencing them, for example using the Illumina sequencing platform. Bignell's tags include a sample identifier so that every nucleic acid can be associated with the correct sample. For instance, Bignell teaches that "[t]he invention relates to methods for indexing samples during the sequencing of polynucleotide templates, resulting in the attachment of tags specific to the source of each nucleic acid sample such that after a sequencing run, both the source and sequence of each polynucleotide can be determined." EX1013(Bignell), 1:24-28.

234. Bignell teaches that this method is that is useful for Next-Gen sequencing techniques. For instance, Bignell explains that the "[m]ethods of the invention are especially useful for the parallel sequencing of a large number of relative[ly] small samples in the same experiment." EX1013(Bignell), 8:21-23. This type of experiment is referred to as multiplex sequencing, in which nucleic acids from multiple samples are sequenced simultaneously. Bignell teaches that

sequencing can be performed using the "Illumina Genome Analyser." EX1013(Bignell), 45:48-50; *also* 43:48-50 ("Solexa/Illumina cluster station").

235. Bignell teaches that its method can be used to sequence biological samples, such as DNA. For instance, Bignell explains that "[i]n a particular embodiment, the primary polynucleotide molecules are DNA molecules." EX1013(Bignell), 8:4-5. Bignell teaches that its method can be used to analyze DNA in "complex genomes (e.g., human genomes)" as well as "less complex genomes, such as those of bacteria, viruses, mitochondria, and the like." *Id.*, 1:29-32; *see also id.*, 8:15-16 ("human genomic DNA molecules"), 8:47-53.

236. Bignell explains that its method adds unique tags to samples of target nucleic acid molecules from different sources. Bignell teaches "target nucleic acids from a plurality of sources can be tagged with a unique, identifiable sequence of nucleic acid bases." EX1013(Bignell), 6:36-40. Bignell refers to the unique tag sequence as a "sample specific tag." *Id.*, 9:48-52. Bignell explains that "a sample specific tag is a nucleic acid sequence tag which is attached to each of the target molecules of a particular sample, the presence of which is indicative of or is used to identify the sample or source from which the target molecules were isolated." *Id.* Further, "[t]he tag is not itself formed by part of the target nucleic acid molecule." *Id.*, 9:44-46. Bignell teaches that the tag is generally "a synthetic

sequence of nucleotides which is added to the target as part of the template preparation step." *Id.*, 9:46-48.

237. Bignell teaches that the "sample specific tag sequence" can be incorporated with the target nucleic acid during the library preparation step. For instance, Bignell teaches that the sample specific tag sequence can be a part of the adaptor. EX1013(Bignell), 9:62-65, 4:40-48 ("sample specific tagged adaptor"). Bignell explains that the tags are incorporated with the target nucleic acids by ligating the tagged adaptor to the target nucleic acid. *Id.*, 4:40-42, 5:5-7, 9:59-62; *see also id.*, 12:38-40 ("Ligation methods are known in the art and utilize standard methods")..

238. The sample specific tags can be incorporated with the target nucleic acid before the target nucleic acid molecules are amplified. EX1013(Bignell), 4:40-53, 5:5-20. The sample tags are also amplified, so they can be analyzed to determine "the sample or source from which the target molecules were isolated." *Id.*, 9:48-52. Because different samples have different tags, Bignell teaches that the "sample specific tag sequence differentiates amplified nucleic acids originating from different samples." *Id.*, 4:10-18. Following amplification, the tagged nucleic acid molecules are pooled with molecules from other samples, and sequenced. *Id.*, 4:19-29. The sequence reads can then be associated with a particular sample based on the sample specific tag sequence. *Id.*, 3:56-59, 6:40-43. Bignell demonstrates

in an example that "the tag sequence could unambiguously determine the source of the sample." *Id.*, 48:11-13.

2. Overview of McCloskey

239. McCloskey describes a method of preparing samples by adding tags to each nucleic acid molecule in the sample. McCloskey explains that the tags allow its method to identify and correct errors and bias stemming from amplification, including "skewing." *See, e.g.*, [0005], [0021], [0106]. For instance, McCloskey explains that PCR amplification can "result in a skewed yield of PCR products such that there is a high degree of redundancy for a small portion of the original genome sequence." EX1012(McCloskey), [0003]. McCloskey teaches that, "[i]n response to the challenges of PCR redundancy and contamination associated with PCR amplification of limited amounts of DNA template, genomic DNA fragments were labeled with molecular sequence barcodes and 'batch-stamps' prior to PCR amplification." *Id.*, [0093]; *see also id.*, [0005], [0021], [0027]. Additionally, McCloskey teaches that its method "also allows for the analysis of mutations arising during PCR amplification." *Id.*, [0106]; *also id.*, [0063] ("PCR errors").

240. McCloskey refers to its tags as "bar-coded oligonucleotides." McCloskey explains that each "bar-coded oligonucleotide" can have a "sequence providing a random barcode." EX1012(McCloskey), [0021]. McCloskey teaches

that "[t]he term 'random barcode' refers to an arbitrary sequence that can uniquely identify a target nucleic acid in an experiment." *Id.*

241. McCloskey also explains that each "random barcode" is unique to a target nucleic acid molecule and that the length of the barcode sequence may be increased to ensure that each barcode sequence is unique. For instance, McCloskey explains that "[s]ome applications will require a larger pool of random-sequence barcodes if more independently derived sequences are required." EX1012(McCloskey), [0101]. McCloskey demonstrates that random barcode sequences with 13 nucleotides can be used which provide about "$1.6x10^6$ different codes." *Id.* McCloskey teaches that the sequences of random barcodes can be even longer—up to 30 nucleotides. *Id.*, [0021]. This would yield over $1x10^{18}$ possible barcodes and would further increase the probability that every nucleic acid molecule is associated with a unique barcode.

242. McCloskey teaches that in addition to a "random barcode" sequence, each "bar-coded oligonucleotide" can also include "a predetermined nucleic acid sequence providing a batch-stamp or experiment verification information." EX1012(McCloskey), [0027]. McCloskey explains that "a batch-stamp may identify a sample, a patient, and/or an analysis date." *Id.* A person of ordinary skill in the art would therefore have understood that a "batch-stamp" includes a sample tag.

243. McCloskey teaches that the tags (*i.e.*, "bar-coded oligonucleotides") are added to the target nucleic acid molecules prior to amplification. For instance, McCloskey explains that "nucleic acid molecules (e.g., genomic DNA fragments) are labeled with distinct sequence tags prior to PCR amplification." EX1012(McCloskey), [0005]; *also id.*, [0033] ("the bar-coded target nucleic acid molecule is amplified"). McCloskey teaches that doing so permits "correcting for skewing that can arise from PCR amplification or the cloning of products." *Id.*, [0005]. McCloskey also teaches that its method "allows for the analysis of mutations arising during PCR amplification," *id.*, [0106], as well as "PCR error," *id.*, [0063]. Following amplification, the tagged nucleic acid molecules are sequenced, and McCloskey provides an example of such sequencing results. *Id.*, [0034], [0063].

244. McCloskey teaches that, in one embodiment of the invention, the tag sequences are on a "hairpin linker" which is added to the target nucleic acid molecules. *See, e.g.*, McCloskey, [0010]; *also id.*, [0092] ("genomic DNA fragments were labelled with molecular sequence barcodes and 'batch-stamps' prior to PCR amplification by including these molecular labels in the hairpin linker sequence (FIG. 3) that is used in hairpin-bisulfite PCR"). A person of ordinary skill in the art would have recognized that McCloskey's use of the hairpin structure was a design choice and would have understood that McCloskey's methods could

also be used with a traditional oligonucleotide design where the random barcode is on a linear structure.

245. McCloskey explains that its method can be used in sequencing DNA from biological samples. For instance, McCloskey explains that its invention may be used for biomedical purposes. EX1012(McCloskey), [0006]. McCloskey also teaches that its invention is "useful in diagnosis and prognosis of some cancers" and other human diseases, including "chronic myelogenous leukemia and scleroderma." *Id.*, [0007]; *see also id.*, [0103] ("fragile X syndrome"). McCloskey also explains that the samples of target nucleic acid molecules can be "clinical samples," *id.*, [0019], and that sample tags ("batch-stamps") can be used to identify "a patient," *id.*, [0027]. McCloskey further teaches that its method can be used to analyze biological samples, such as "[h]uman genomic DNA [that] was isolated from blood." *Id.*, [0045].

246. McCloskey further explains that "the barcoding method thus provides a definitive solution to the problem identified previously . . . in which multiple amplified sequences are derived from a single original molecule when template DNA is limited in amount or of poor quality." EX1012(McCloskey), [0106]. The method described by McCloskey would be particularly useful in sequencing experiments involving a small quantity of DNA, circulating DNA in blood

samples, or other preserved and/or degraded clinical samples in which the amount of DNA might be compromised.

3. A Person of Ordinary Skill in the Art Would Have Modified Bignell to Incorporate Elements Shown in McCloskey

247. A person of ordinary skill in the art would have considered making modifications to the methods described by McCloskey based on Bignell.

248. McCloskey teaches methods of preparing a sample for sequencing which allows for the individual nucleic acid molecules to be identified and traced during the sequencing workflow. EX1012(McCloskey), [0002], [0106]. McCloskey, however, does not describe techniques for the sequencing step of the nucleic acid workflow. Those techniques were known in the art, as McCloskey notes. *Id.*, [0034] ("[m]ethods… for… sequencing are standard in the art"). Therefore, a person of ordinary skill in the art who was interested in implementing McCloskey's method would have looked to other references describing known sequencing technologies.

249. One reference describing sequencing technologies was Bignell. Bignell provides a method for sequencing using Next-Gen sequencing techniques. EX1013(Bignell), 17:12-27, 41:51-61, 46:45-47:3, 6:36-43. For instance, Bignell explains that the "[m]ethods of the invention are especially useful for the parallel sequencing of a large number of relative[ly] small samples in the same

experiment." *Id.*, 8:21-23. Bignell explains how sequencing can be performed using the "Illumina Genome Analyser." *Id.*, 45:48-50.

250. A person of ordinary skill in the art would have been motivated to combine McCloskey specifically with a Next-Gen sequencing technique, such as the ones described in Bignell. As I explained above, by early 2009, Next-Gen sequencing technologies were widely adopted because they provided sequencing methods which were both cost-effective and highly efficient. *See, e.g.*, EX1014, 1135. Next-Gen sequencing technologies have been praised for "undeniably reinvigorating this field" and creating "the incentive for developing entirely new strategies for DNA sequencing." *Id.* It made sequencing more affordable and available for all types of experiments and analyses. Thus, a person of ordinary skill in the art would have been motivated to implement a Next-Gen sequencing technique, such as the technique described by Bignell, with McCloskey's method.

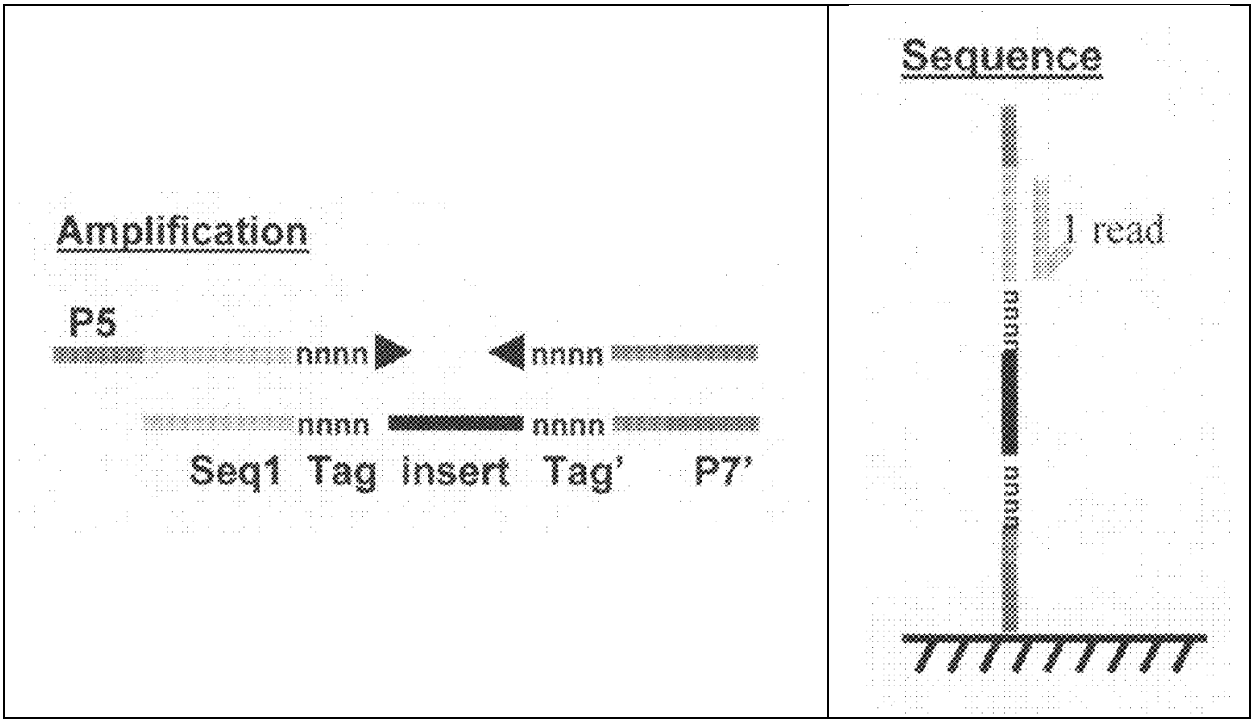
251. Bignell's method includes a number of amplification steps. As I explained above, a person of ordinary skill in the art would have known that the amplification process typically introduced bias or errors. *See* EX1012(McCloskey), [0003]; EX1025, 8966; EX1018, [0002]. Such bias could come from making more copies of some nucleic acids than others. Errors could arise when there were mutations that occurred during the copying or synthesis processes. McCloskey teaches that its method of adding unique barcodes to each

nucleic acid molecule provides a way to identify and correct for errors and/or bias introduced by amplification. EX1012(McCloskey), [0005]. A person of ordinary skill in the art would have therefore recognized that incorporating McCloskey's unique barcodes with Bignell's method of Next-Gen sequencing would have provided a solution to errors and/or bias caused by the amplification steps.

252. Further, a person of ordinary skill in the art would have recognized that McCloskey and Bignell describe analogous and compatible tagging techniques. Both McCloskey and Bignell explain that adding sample tags to nucleic acid molecules provides the advantage of being able to identify and track the original sample of the nucleic acid molecules. EX1012(McCloskey), [0027]; EX1013(Bignell), 1:24-28, 3:54-59. Additionally, both McCloskey and Bignell teach that the sample tags are incorporated with the nucleic acid molecules before they are amplified and then used in subsequent steps. EX1012(McCloskey), [0005], [0009]; EX1013(Bignell), 4:40-53, 5:5-20. They also both teach that the sample tags can be incorporated with adaptors. EX1012(McCloskey), [0009] ("third sequence"); EX1013(Bignell), 9:62-65, 4:40-48. Further, both describe methods for the nucleic acid sequencing workflow using DNA samples. EX1012(McCloskey), [0005], [0045]; EX1013(Bignell), 7:25-26, 8:4-5.

253. A person of ordinary skill in the art would have had a reasonable expectation of success in combining the "random barcode sequence" described by

McCloskey on the adaptor described by Bignell. Bignell explains that the adaptors may serve different functions and that "[a]dditional sequence elements may be included." EX1013(Bignell), 11:65-12:9; *also id.*, 9:62-67. Bignell teaches that the sequence of the adaptors "may be selected by the user" based on the desired function. *Id.* Bignell also shows that multiple tags can be added to a target DNA molecule without compromising the read length of the target nucleic acid, because a sequencer can read the target nucleic acid and tag sequences separately.



EX1013(Bignell), Fig. 3; *also id.*, 17:36-39, 6:61-7:3.

254. McCloskey teaches that a tag can include both a "random barcode" sequence and a "batch-stamp" sequence, which refers to a sample specific sequence. EX1012(McCloskey), [0046], [0027]. Therefore, a person of ordinary

skill in the art would have understood that the adaptors described in Bignell could be modified to additionally include a "random barcode" sequence, like the ones described by McCloskey.

255. A person of ordinary skill in the art would have also been familiar with techniques for creating and attaching unique identifiers to nucleic acid molecules in a sample, as those techniques were well-known. EX1010(Chee), 5:31-36; EX1019(Brenner), 3:36-65, 7:18-35, 10:41-49. Thus, a person of ordinary skill in the art could have employed known techniques to modify Bignell to use tags that included sample identifiers (as taught by both Bignell and McCloskey) and also McCloskey's unique barcode.

256. In the mid-2000s, multiple people invented techniques for using tags to track individual nucleic acids in a sample. For example, both Chee and the '730 patent show that different people independently developed the same improvement to the prior art to solve a known problem. Chee invented techniques for using unique identifiers to correct for errors and bias at least a few weeks or months before the '730 patent's provisional applications were filed. That shows that those in the art at that time would have been motivated to combine McCloskey's tags that include unique barcodes with the tagging methods described in Bignell.

257. Therefore, a person of ordinary skill in the art would have been motivated to combine McCloskey's "random barcode" sequence with Bignell's

Next-Gen sequencing technique and would have reasonably expected for the combination of McCloskey and Bignell to work.

4. Bignell and McCloskey Describe the Elements of Independent Claims 1 and 9 of the '730 Patent

258. It is my opinion that Bignell in combination with McCloskey renders claims 1 and 9 obvious. Claim 1 (EX1001) is exemplary and reproduced below.

> 1. A method for correcting for errors or bias introduced during nucleic acid analysis workflow, the method comprising the steps of:
>
> obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin;
>
> introducing a set of differentiator tags, wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag;
>
> amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons;
>
> sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence; and
>
> correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count.

a) Preamble

259. The preamble of claims 1 and 9 specify "[*a*] *method for correcting for errors or bias introduced during nucleic acid analysis workflow*."

260. Bignell discloses a method for a nucleic acid workflow. As I explained above, Bignell teaches a method of sequencing nucleic acid molecules, such as DNA, using Next-Gen sequencing techniques. *See, e.g.*, EX1013(Bignell), 17:12-27, 45:48-50. Bignell teaches a method of using unique sample tags to facilitate "multiplex sequencing."

261. McCloskey discloses a method for correcting for error or bias introduced during a nucleic acid analysis workflow. EX1012(McCloskey), [0002]. As I explained above, McCloskey teaches that the amplification step in the nucleic acid sequencing workflow can introduce errors and bias. For instance, McCloskey explains that PCR amplification can "result in a skewed yield of PCR products such that there is a high degree of redundancy for a small portion of the original genome sequence." *Id.*, [0003].

262. McCloskey explains that its methods can correct for errors and bias introduced during the amplification process. McCloskey teaches that "[i]n response to the challenges of PCR redundancy and contamination associated with PCR amplification of limited amounts of DNA template, genomic DNA fragments were labeled with molecular sequence barcodes and 'batch-stamps' prior to PCR amplification." EX1012(McCloskey), [0093]. McCloskey explains that "[b]arcoding permits . . . correcting for skewing that can arise from PCR amplification or the cloning of products." *Id.*, [0005]. Additionally, McCloskey

teaches that its method "allows for the analysis of mutations arising during PCR amplification," *id.*, [0106], as well as "PCR error," *id.*, [0063].

b) "obtaining a biological sample comprising a plurality of target nucleic acid molecules from more than one locus of origin"

263. Claims 1 and 9 specify the above limitation, which Bignell and McCloskey teach.

264. Bignell teaches that its method can be used to analyze biological samples, such as DNA molecules. For instance, Bignell explains that "[i]n a particular embodiment, the primary polynucleotide molecules are DNA molecules," such as "human genomic DNA molecules." EX1013(Bignell), 8:4-16. Bignell's examples also use DNA from "human samples." 26:10-13.

265. Bignell also shows that the sample comprises target nucleic acids from "*more than one locus of origin*." For instance, Bignell explains that its methods can be used to analyze "human genomes," as well as "less complex genomes, such as those of bacteria, viruses, mitochondria, and the like." EX1013(Bignell), 1:29-32. Bignell explains that samples can include DNA "molecules [that] represent the entire genetic complement of an organism, and are genomic DNA molecules." *Id.*, 8:5-10. Bignell also teaches that its methods can be used with "particular sub-sets of polynucleotide sequences or genomic DNA, such as… particular chromosomes." *Id.*, 8:10-13. Thus, a person of ordinary skill

in the art would have understood that analyzing genomes implicates the analysis of nucleic acid molecules from multiple loci.

266. McCloskey also teaches its method involves "*obtaining a biological sample*," such as "naturally-occurring" DNA molecules. EX1012(McCloskey), [0018]. For instance, McCloskey teaches that its method used "DNA samples [which] were obtained from blood" to analyze the "[h]uman genomic DNA [that] was isolated from blood." *Id.*, [0045]. McCloskey also demonstrates an example where the sample is taken from the "FMR1 locus" on the genome. *Id.*, [0043].

267. Further, McCloskey teaches that the sample comprises target nucleic acids from "*more than one locus of origin*." For instance, McCloskey teaches that its methods can be used to analyze "genomic DNA fragments," which refers to DNA fragments which exist at any location on the genome. EX1012(McCloskey), [0005]. McCloskey explains that its invention may be used for biomedical purposes or forensic purposes. *Id.*, [0006]. McCloskey also explains that its invention is useful for the "diagnosis and prognosis of some cancers" and other human diseases, including "chronic myelogenous leukemia and scleroderma." *Id.*, [0007]; *see also id.*, [0103] ("fragile X syndrome"). McCloskey also explains that the samples of target nucleic acid molecules can be "clinical samples," *id.*, [0019], and that sample tags ("batch-stamps") can be used to identify "a patient," *id.*, [0027]. As I explained above, a person of ordinary skill in the art would have

understood that using McCloskey's methods for these purposes implicate the analysis of target nucleic acids from multiple loci.

c) Claim 1: "introducing a set of differentiator tags wherein members of said set of differentiator tags are associated with members of said plurality, such that one or more of said loci of origin are associated with more than one differentiator tag"

268. Claim 1 specifies the above limitation, which Bignell and McCloskey together teach.

269. Bignell describes a method where tags are added to nucleic acids in a sample. For instance, Bignell teaches "target nucleic acids from a plurality of sources can be tagged with a unique, identifiable sequence of nucleic acid bases." EX1013(Bignell), 6:36-40; *see id.*, 7:59-64. Bignell teaches that the tag is generally "a synthetic sequence of nucleotides which is added to the target as part of the template preparation step." *Id.*, 9:46-48. Further, Bignell explains that "[a] nucleic acid sequence tag characteristic of the source is attached to each of the target molecules in each isolated sample before the sample is immobilized for sequencing." *Id.*, 9:41-44.

270. Bignell also teaches that these tags allow for the nucleic acids in one sample to be differentiated from the nucleic acids in another sample. For instance, Bignell explains that "a sample specific tag is a nucleic acid sequence tag which is attached to each of the target molecules of a particular sample, the presence of

which is indicative of or is used to identify the sample or source from which the target molecules were isolated." EX1013(Bignell), 9:48-52; *see also id.*, 13:16-22, 13:32-35, 21:18-22.

271. Finally, Bignell teaches that the tag sequence can be incorporated in the adaptor. EX1013(Bignell), 9:62-65 ("tags are on the adaptor"), 4:40-48 ("sample specific tagged adaptor"), 9:59-10:4, 10:23-26. Bignell also teaches that the sequence of the adaptors "may be selected by the user such that the desired sequence elements are ultimately included." *Id.*, 11:65-12:9.

272. As I explained above in § VII.D.3, a person of ordinary skill in the art would have modified the adaptors in Bignell to incorporate the molecule-specific unique barcodes in McCloskey.

273. McCloskey teaches a method in which tags are introduced and associated with each nucleic acid molecule in a sample. *See, e.g.*, EX1012(McCloskey), [0005], [0009]. For instance, McCloskey describes a method of synthesizing and adding (*i.e.,* introducing) "bar-coded oligonucleotides" to a sample of nucleic acids. *Id.*, [0046]; *see also id.*, [0047] ("[t]hese steps allow for the extension of the annealed oligo onto the genomic template, resulting in a bar-coded template for subsequent amplification").

274. McCloskey refers to the tags as "bar-coded oligonucleotides." EX1012(McCloskey), [0021]. McCloskey's barcodes can be comprised of four

parts. These four parts include: a "first sequence" that is an adaptors or primer, *id.*, [0020], [0022], a "second sequence" called a "random barcode" that will "uniquely identify a target nucleic acid in an experiment," *id.*, [0021], a "third sequence" that is another adaptor or primer, *id.*, [0020], [0022], and a "fourth sequence" called a "batch-stamp" that identifies a sample or a patient, *id.*, [0027].

275. McCloskey teaches that "[t]he term 'random barcode' refers to an arbitrary sequence that can uniquely identify a target nucleic acid in an experiment." EX1012(McCloskey), [0021]. In other words, no two nucleic acids in the starting sample will be associated with the same barcode. McCloskey reports the results of experiments where every nucleic acid was given a unique barcode. *Id.*, [0053], [0063]. Thus, a person of ordinary skill in the art would have understood that "random barcodes" in McCloskey meet the "*differentiator tags*" element of the claims.

276. McCloskey also teaches that "*one or more of said loci of origin are associated with more than one differentiator tag*." McCloskey explains that each "random barcode" is unique to a target nucleic acid molecule and that the length of the barcode sequence may be increased to ensure that each barcode sequence is unique. For instance, McCloskey explains that "[s]ome applications will require a larger pool of random-sequence barcodes if more independently derived sequences are required," EX1012(McCloskey), [0101], and explains that the sequence for the

random barcodes may be up to 30 nucleotides. *Id.*, [0021]. McCloskey teaches that random barcodes with 13 nucleotides can be used which provides about "$1.6x10^6$ different codes" and that "the probability that two of [the cloned PCR products] would be different genomic fragments labeled with identical barcodes is only 0.0031." *Id.*, [0101].

277. McCloskey also teaches that the sequences of random barcodes can be even longer—up to 30 nucleotides. *Id.*, [0021]. This would yield over $1x10^{18}$ possible barcodes and would further increase the probability that every nucleic acid molecule is associated with a unique barcode. Thus, as I explained above, a person of ordinary skill in the art would have understood they could synthesize a longer barcode sequence to further increase the probability that each nucleic acid molecule is associated with a unique barcode.

278. Further, McCloskey teaches that its random barcodes provides "a unique identity to each target nucleic acid molecule in the sample prior to amplification." EX1012(McCloskey), [0021]. Consequently, a random barcode "can uniquely identify a target nucleic acid." *Id.* McCloskey demonstrates this is its examples by showing that multiple nucleic acid molecules "had distinct barcode regions," *id.*, [0053], even though the molecules all originated from the same location on the genome: the FMR1 locus, *id.*, [0043]. *See also id.*, [0063] (obtaining multiple sequences, "each of which had a unique barcode"). Therefore,

McCloskey teaches that "*one or more of said loci of origin are associated with more than one differentiator tag*."

d) Claim 9: "introducing a set of unique differentiator tags wherein members of said set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules, such that any given tag is uniquely associated with said locus of origin"

279. Claim 9 specifies the above limitation, which Bignell and McCloskey together teach. During prosecution, Applicant made clear that "*differentiator tags*" must be unique. Therefore, including the word "*unique*" in this claim therefore does not add meaning to the term "*differentiator tag*." McCloskey's unique barcodes are "*unique differentiator tags*" for the same reasons they are "*differentiator tags*."

280. This limitation further requires that the "*set of differentiator tags are randomly associated with members of said plurality of target nucleic acid molecules*." McCloskey's random barcodes are "*randomly associated*" with target nucleic acid molecules. *See, e.g.*, EX1012(McCloskey), [0017]. McCloskey's method involves synthesizing the oligonucleotides for the random barcodes, *id.*, 0021], and then introducing them into a sample to attach to the target nucleic acid molecules, *id.*, [0017], [0046]-[0047]. Further, McCloskey's "random barcodes" do not have a region that specifically targets a nucleic acid, and each random barcode is able to attach to any nucleic acid in the sample. McCloskey's "random

barcodes" are allowed to randomly associate with nucleic acids when added to a sample. Therefore, a person of ordinary skill in the art would have understood that the tags in McCloskey meet this claim element.

281. This limitation also further requires that "*any given tag is uniquely associated with said locus of origin*." McCloskey teaches that "[t]he term 'random barcode' refers to an arbitrary sequence that can uniquely identify a target nucleic acid in an experiment." EX1012(McCloskey), [0021]. This means that the random barcodes described by McCloskey are uniquely associated with a target nucleic acid, so that every target nucleic acid will be uniquely associated with a random barcode. Each nucleic acid is associated with one locus, and therefore, any tag is uniquely associated with one locus of origin.

e) "amplifying each of the plurality of tagged target nucleic acid molecules to generate amplicons"

282. Claims 1 and 9 specify the above limitation, which Bignell and McCloskey teach.

283. Bignell teaches that the tagged target nucleic acid molecules are amplified to generate amplicons. For instance, Bignell explains that the "tagged adaptor" sequence is added to the ends of the target nucleic acid molecule to form a sample of "adaptor-target-adaptors." EX1013(Bignell), 4:40-48. The sample of "adaptor-target-adaptors" corresponds to "*tagged target nucleic acid molecules*."

Then the sample(s) of the "adaptor-target-adaptors" are amplified. *Id.*, 5:5-22; *see also id.*, 4:49-53, 12:65-13:15, 15:5-11, 15:14-15. Amplification "generate[s] amplified nucleic acids from each sample." *Id.*, 5:17-20; *see also id.*, 4:51-53; 10:16-19 ("amplified copies").

284. McCloskey also teaches that the tagged target nucleic acid molecules are amplified to generate amplicons. For instance, McCloskey describes "amplifying the bar-coded target nucleic acid molecule… to produce an amplification product." EX1012(McCloskey), [0032]; *see id.*, [0031], [0033], [0035].

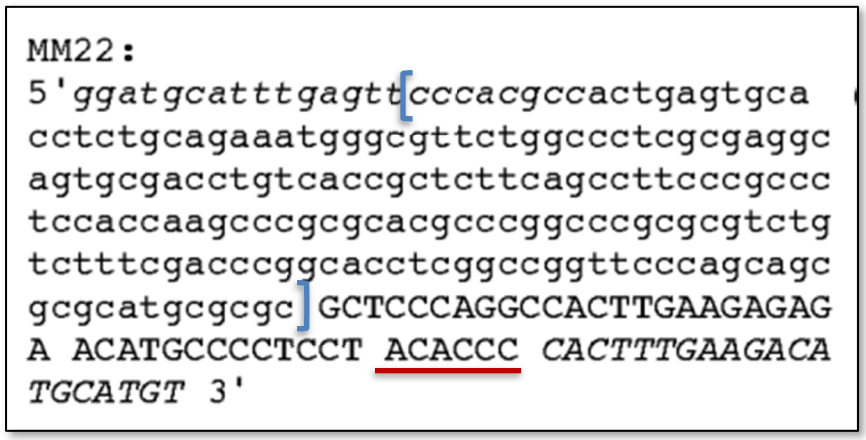
f) "sequencing the amplicons obtained in said amplifying step to obtain sequence reads of each of the amplicons, wherein each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence"

285. Claims 1 and 9 specify the above limitation, which Bignell and McCloskey together teach.

286. Bignell teaches that its method involves a sequencing step, in which the both the target nucleic acid molecule sequence and the tag sequence are read. EX1013(Bignell), 6:61-7:3, 4:56-62, 5:25-30, 17:4-11, 17:36-39. For instance, Bignell's methods allow for the preparation of "a nucleic acid sample for a sequencing reaction wherein both the target sequence and the tag sequence can be determined." EX1013(Bignell), 3:56-59. Bignell explains that the construct can

be prepared so that "the tag sequence is adjacent to the target sequence, and the tag and target are read in a single read." *Id.*, 6:66-7:3. Bignell also describes a process where the tag and target can be read in separate reads. *Id.*, 6:61-66.

287. McCloskey also teaches a method of "*sequencing the amplicons*," in which both the target nucleic acid molecule sequence and the tag sequence are read. *See, e.g.*, EX1012(McCloskey), [0063]. McCloskey teaches that the "amplification products" can be sequenced, and that sequencing methods "are known in the art." *Id.*, [0034]. McCloskey provides an example of a sequencing of a "representative amplified product, MM22." *Id.*, [0051] (annotated below).



288. McCloskey explains that "[t]he sequences [of MM22] in upper case corresponds to the sequence of the bar-coded oligonucleotide," including "the 5 nucleotides corresponding to the barcode (5' ACACCC 3')." EX1012(McCloskey), [0052]. I have annotated the sequence above to underline the barcode sequence in red.

289. McCloskey explains that "[t]he sequences [of MM22] in lower case is the region complementary to FMR1 that was not represented in the bar-coded oligonucleotide, starting with the rightward primer binding site (italicized.)" EX1012(McCloskey), [0052]. I have annotated the sequence above to include blue brackets around this region. A person of ordinary skill in the art would therefore have understood that the lower case, non-italicized region is the region complementary to FMR1 (*i.e.*, the target nucleic acid molecule). A person of ordinary skill in the art would further have understood that a sequence read of the complementary region also provides the sequence read of the target nucleic acid molecule.

290. Thus, Bignell and McCloskey's combined method involves sequencing the amplified products where "*each of the sequence reads comprises a target nucleic acid molecule sequence and a differentiator tag sequence*."

g) Claim 1: "correcting for error or bias introduced during said workflow by collapsing target:differentiator tag combinations observed more than once into a single count"

291. Claim 1 specifies the above limitation, which Bignell and McCloskey together teach.

292. As I explained above for the claim preamble, McCloskey's method can both detect and correct for errors and bias caused by the amplification and sequencing processes. *See* EX1012(McCloskey), [0002].

293. McCloskey teaches that using unique barcodes attached to each molecule in the starting sample allows for "correcting for skewing that can arise from PCR amplification." EX1012(McCloskey), [0005]. A skilled person would have understood that McCloskey corrects that skewing by determining that all molecules with the same barcode correspond to the same starting molecule in the sample. McCloskey states that identifying duplicate barcodes allows it to determine that there are "multiple amplified sequences derived from a single original molecule." *Id.*, [0093]. In McCloskey's examples, McCloskey uses the barcodes to ensure that it does not double count duplicate sequences in its analysis. *Id.*, [0053], [0063].

294. McCloskey's barcodes allow it "to identify… redundant sequences and to quantify accurately the proportion of cells carrying a particular sequence variant by ***counting only distinctly tagged sequences***." EX1012(McCloskey), [0093]; *see id.*, [0106] (can "quantif[y] the relative abundance" of molecules). McCloskey explains that the "ability to distinguish redundant from valid data is an immensely powerful tool." *Id.*, [0053]. This process corrects for any bias introduced by uneven amplification of the sample.

295. McCloskey's method of identifying duplicates to prevent double counting is the same technique that the '730 patent describes as "collapsing" differentiator tags and targets. EX1001, 28:32-44. The '730 patent shows

performing the "collapsing" step by counting only unique tags to determine the number of times a sequence appeared in the original sample. *Id.*, 30:48-52, 72:63-73:2. The '730 patent shows that duplicates of a tag are identified but not included in the count. *Id.* McCloskey shows doing the same thing and therefore teaches "*collapsing*" duplicate "*target:differentiator tag combinations… into a single count*."

h) Claim 9: "correcting for error or bias introduced during said workflow by recognizing target/differentiator tags having the same sequence as being derived from the same input molecule"

296. Claim 9 specifies the above limitation, which Bignell and McCloskey together teach. McCloskey describes a method of "*correcting for error or bias introduced during said workflow*," as I discussed above with respect to claim 1.

297. This limitation is similar to claim 1, but the limitation differs slightly in that it requires "*recognizing target/differentiator tags having the same sequence as being derived from the same input molecule*."

298. McCloskey teaches that using unique barcodes attached to each molecule in the starting sample allows for "correcting for skewing that can arise from PCR amplification." EX1012(McCloskey), [0005]. McCloskey's barcodes allow it "to identify… redundant sequences and to quantify accurately the proportion of cells carrying a particular sequence variant by ***counting only***

distinctly tagged sequences." EX1012(McCloskey), [0093]; *see id.*, [0106] (can "quantif[y] the relative abundance" of molecules). McCloskey explains that the "ability to distinguish redundant from valid data is an immensely powerful tool." *Id.*, [0053]. This process corrects for any bias introduced by uneven amplification of the sample.

299. A skilled person would have also understood that McCloskey corrects that skewing by determining that all molecules with the same barcode correspond to the same starting molecule in the sample. McCloskey states that identifying duplicate barcodes allows it to determines that there are "multiple amplified sequences derived from a single original molecule." *Id.*, [0093]. McCloskey therefore teaches "recognizing" random barcodes that have the same unique sequence as "*being derived from the same input molecule*."

5. Claim 4

300. Claim 4 depends on claim 1 and further recites "*wherein said biological sample is selected from the group consisting of blood and biopsy tissue*."

301. McCloskey teaches this limitation. For instance, McCloskey explains its method can be used to analyze "[h]uman genomic DNA [that] was isolated from blood." EX1012(McCloskey), [0045].

6. Claim 5

302. Claim 5 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules comprises bacterial or viral nucleic acids isolated from said biological sample*."

303. Bignell teaches this limitation. For instance, Bignell explains its method can be used to analyze "less complex genomes, such as those of bacteria, viruses, mitochondria, and the like." EX1013(Bignell), 1:29-32; *also id.*, 8:47-53.

7. Claim 6

304. Claim 6 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules*."

305. McCloskey teaches this limitation. McCloskey teaches that its method can be used to analyze biological samples, such as "[h]uman genomic DNA [that] was isolated from blood." EX1012(McCloskey), [0045]. As I explained above, tumors cells shed DNA into the bloodstream which can be referred to as "circulating tumor DNA." McCloskey also teaches that its invention is "useful in diagnosis and prognosis of some cancers" and it "provides methods to assess the variations of methylation, and their quantification, in a heterogeneous population of cancer cells and normal cells." *Id.*, [0007]. A person of ordinary skill in the art would have understood that McCloskey's methods for analyzing blood also involve analyzing circulating tumor DNA.

306. McCloskey further explains that "the barcoding method thus provides a definitive solution to the problem identified previously . . . in which multiple amplified sequences are derived from a single original molecule when template DNA is limited in amount or of poor quality." EX1012(McCloskey), [0106]. A person of ordinary skill in the art would have understood that the method described by McCloskey would be particularly useful in sequencing experiments involving a small quantity of DNA, such as circulating tumor DNA.

E. Bignell, McCloskey, and Bentley Render Obvious Claim 2

1. Overview of Bentley

307. Bentley describes the advances in sequencing technology and how they can be used to analyze whole genomes. EX1017(Bentley), 545. Bentley also provides overviews of "[n]ew sequencing technologies," including sequencing by synthesis. *Id.*, 546. Two available sequencing by synthesis systems were Illumina's Solexa system and the Roche 454 system. *Id.*, 546, 548.

308. Bentley explains that Next-Gen sequencing, such as Illumina's Solexa system, can be used to generate a consensus sequence. *Id.*, 549 ("multiple reads from the same sample can be aligned and used to derive a consensus base-call at each position"), 550 (Fig. 2). Bentley also explains that "errors" are "excluded from being true base calls" and therefore are not included in the consensus

sequence. Thus, the generation of a consensus sequence recognizes and corrects for error.

2. A Person of Ordinary Skill in the Art Would Have Configured Bignell and McCloskey to Generate a Consensus Base Call as Taught by Bentley

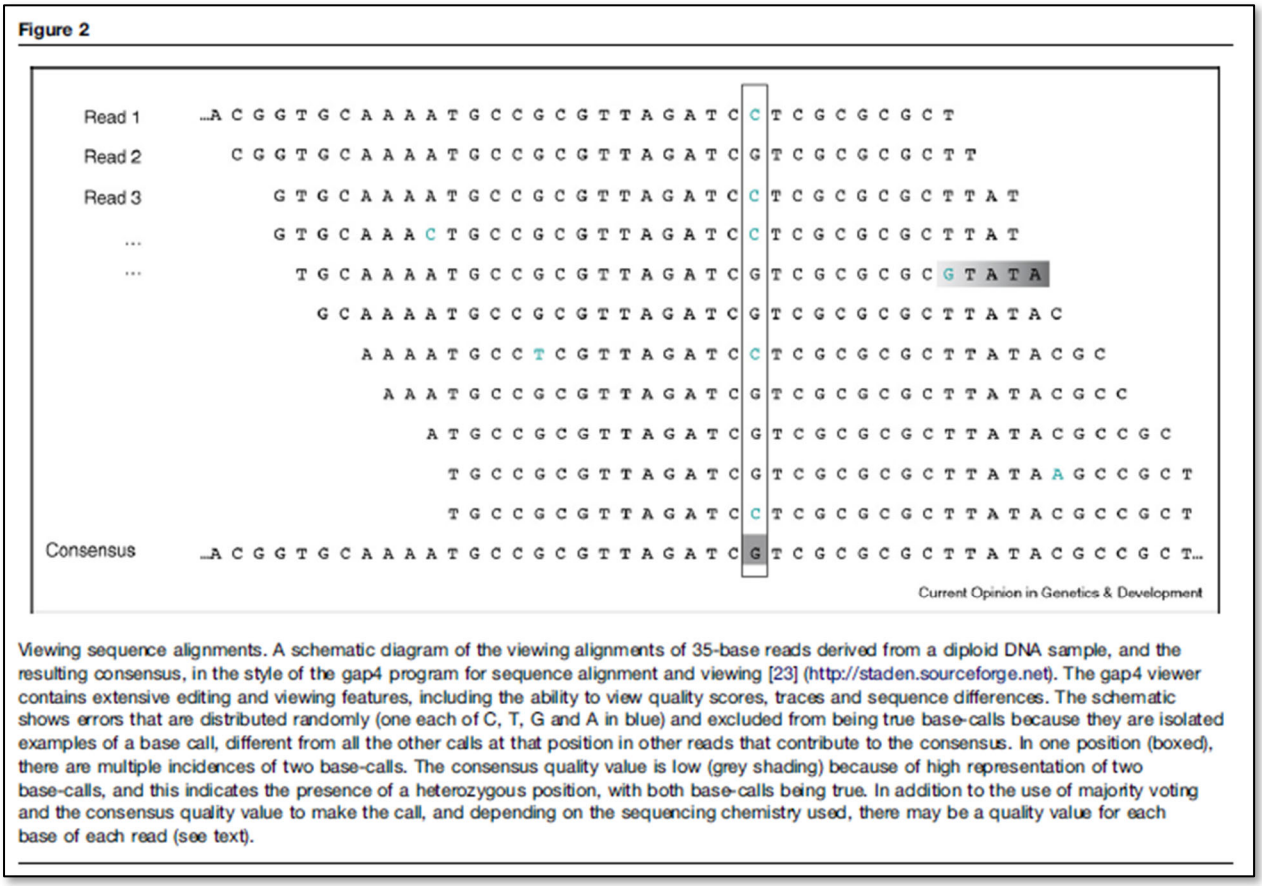
309. A person of ordinary skill in the art would have been motivated to practice McCloskey and Bignell based on Bentley's teachings.

310. McCloskey includes examples where errors were identified and corrected. For instance, McCloskey explains that "[e]ach sequence was proofread against the sequence trace" and that "errant base calling was corrected manually before being presented here." EX1012(McCloskey), [0098]. A person of ordinary skill in the art would have understood that checking for errors improves the accuracy of sequencing, and that conventional knowledge is reflected in McCloskey, describes a process for identifying "errant base calling." A skilled person would have been motivated to look to other references that describe details about conventional techniques for identifying and correcting errors.

311. One such reference is Bentley, which describes a method to recognize and correct errors that arise during the nucleic acid sequencing workflow, by generating a consensus base call or a consensus sequence. Bentley provides an overview of this well-known technique for analyzing sequence reads and explains

that a consensus base call can be generated with data from Next-Gen and traditional Sanger sequencing technologies.

312. Figure 2, pictured below, demonstrates how the alignment of several sequence reads can be used to generate a consensus sequence using data from Next-Gen sequencing technologies. EX1017(Bentley), 550.



313. Bentley explains that Figure 2 "shows errors that are distributed randomly (one each of C, T, G and A in blue) and excluded from being true base-calls because they are isolated examples of a base call, different from all the other calls at that position in other reads that contribute to the consensus." *Id.* This

means that errors are recognized and are not incorporated in the consensus sequence. Thus, a person of ordinary skill in the art would have been motivated to generate a consensus sequence as taught by Bentley with the sequencing techniques taught by McCloskey and Bignell as a method to identify and correct for error.

314. Bentley also teaches that a consensus sequence can correct for errors that stem from "misincorporation by the polymerase." EX1017(Bentley), 548. A person of ordinary skill in the art would understand that this includes errors that were introduced by the incorrect addition of nucleotides during the synthesis process, such as during PCR amplification. Bentley explains that this type of error will generate "a high-confidence wrong base-call" which generally "will not affect consensus base-calls." *Id.*. Thus, the errors will be recognized as true errors with "high-confidence" and will not be included in the consensus sequence.

315. McCloskey's methods are also concerned about errors introduced during the amplification step. For instance, McCloskey teaches that its method "allows for the analysis of mutations arising during PCR amplification," EX1012(McCloskey), [0106], as well as "PCR error," *id.*, [0063]. A person of ordinary skill in the art would therefore have been motivated to generate a consensus sequence as taught by Bentley with McCloskey's methods.

316. Bentley also teaches that when "each sequencing template is clonal in origin, high-quality base-calls are generated separately for each variant." EX1017(Bentley), 546. A person of ordinary skill in the art would have understood that this means a consensus base call could be generated with copies of the same molecule. Thus, a person of ordinary skill in the art would have had a reasonable expectation of success in generating a "high-quality base-call" for the molecules in McCloskey that have the same random barcode, since those molecules are "clonal in origin."

317. Finally, a person of ordinary skill in the art would have had a reasonable expectation of success in generating a consensus sequence with data from the sequencing techniques for Bignell and McCloskey. Bentley explains that a consensus sequence could be generated with data from Next-Gen sequencing technologies, such as Illumina's Solexa system, as well as data from traditional Sanger sequencing (referred to as "capillary sequencing") using similar methods. Bignell describes techniques that can be used with Illumina's sequencer. A skilled person would have been motivated to configure the combined system to generate consensus sequences when the amplified products contained duplicates (amplified nucleic acids with the same barcode). The ability to create consensus reads was a feature of Illumina's system, and it improved accuracy as Bentley explains.

318. Bentley explains that "multiple reads from the same sample can be aligned and used to derive a consensus base-call at each position." EX1017(Bentley), 549. This type of consensus analysis "provides high-confidence base-calling." *Id.* Accuracy of the consensus is measured using an algorithm called "Phred" which "provides a numerical score (i.e. quality value) that is an estimate of the probability of error for each base-call in a raw read." *Id.*

319. Next-Gen sequencing technologies which use "[b]ase by base sequencing," such as Illumina's Solexa system, also "provide[] information on the raw data collected at each base." EX1017(Bentley), 549. Like for traditional Sanger sequencing, this enables the derivation of "quality values" that estimate the probability of error for each base call. *Id.* Further, Bentley explains that "[c]onsensus scoring and quality values can be used to call SNPs ***in the same way as that developed for capillary sequence data*** (illustrated schematically in Figure 2)." *Id.*

320. Thus, a person of ordinary skill in the art would have configured Bignell and McCloskey to generate a consensus sequence, and would have had a reasonable expectation of success in doing so.

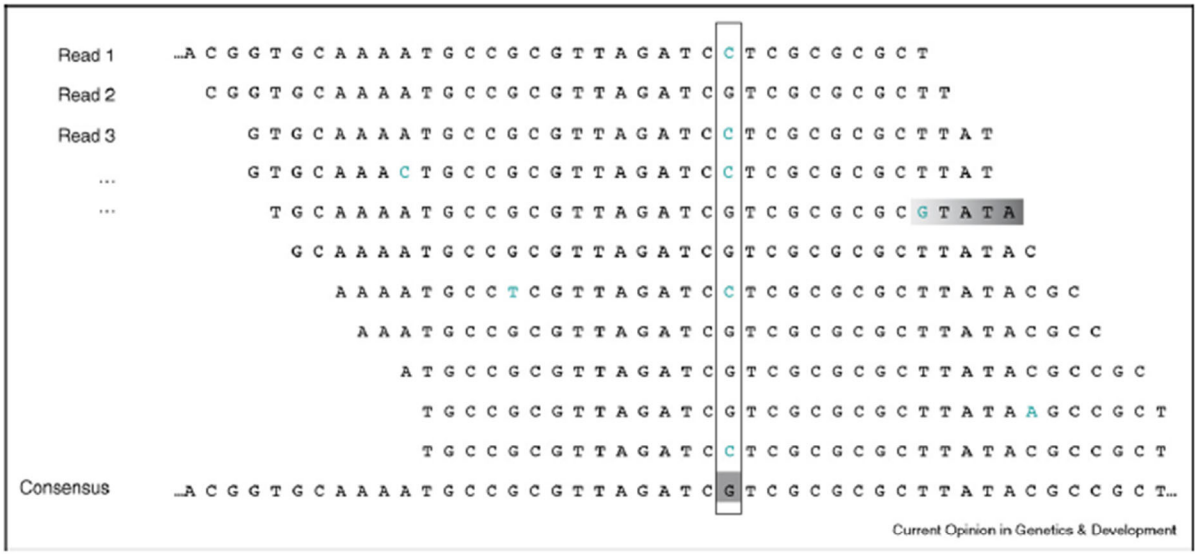
3. Claim 2

321. Claim 2 depends on claim 1 and further recites "*wherein said correcting step further comprises collapsing all sequence reads comprising a same*

differentiator tag into a single read before determining a consensus base call for the target nucleic acid molecule sequences."

322. As I explained above, Bentley teaches "*determining a consensus base call for the target nucleic acid molecule sequences*" which involves "*collapsing*" the sequence reads "*into a single read*." Bentley explains that "multiple reads from the same sample can be aligned and used to derive a consensus base-call at each position," EX1017(Bentley), 549, and illustrates this in Figure 2, below, *id.*, 550.

Figure 2



Viewing sequence alignments. A schematic diagram of the viewing alignments of 35-base reads derived from a diploid DNA sample, and the resulting consensus, in the style of the gap4 program for sequence alignment and viewing [23] (http://staden.sourceforge.net). The gap4 viewer contains extensive editing and viewing features, including the ability to view quality scores, traces and sequence differences. The schematic shows errors that are distributed randomly (one each of C, T, G and A in blue) and excluded from being true base-calls because they are isolated examples of a base call, different from all the other calls at that position in other reads that contribute to the consensus. In one position (boxed), there are multiple incidences of two base-calls. The consensus quality value is low (grey shading) because of high representation of two base-calls, and this indicates the presence of a heterozygous position, with both base-calls being true. In addition to the use of majority voting and the consensus quality value to make the call, and depending on the sequencing chemistry used, there may be a quality value for each base of each read (see text).

Figure 2 shows that 11 reads are collapsed into one "consensus" read. The consensus sequence includes the base call that appears most frequently. Errors are "excluded from being true base-calls" and are therefore not incorporated in the consensus sequence. *Id.*, 550.

323. It would have been obvious to generate a consensus sequence, like this consensus sequence taught by Bentley, for "*all sequence reads comprising a same differentiator tag*." Bentley teaches that when "each sequencing template is clonal in origin, high-quality base-calls are generated separately for each variant." EX1017(Bentley), 546. A person of ordinary skill in the art would have understood that this means a consensus base call (or sequence) could be generated with copies of the same molecule.

324. McCloskey teaches generating and identifying copies of the same molecule by introducing a "random barcode." A random barcode can "uniquely identify a target nucleic acid." EX1012(McCloskey), [0021]. Therefore, a person of ordinary skill in the art would have found it obvious to generate a consensus base call with "*all sequence reads comprising a same differentiator tag*."

4. Claim 3

325. Claim 3 depends on claim 2 and further recites "*further comprising the step of determining the presence of a sequencing or amplification error based upon a number of said differentiator tag sequences*."

326. As I explained above for McCloskey and Bignell, McCloskey uses sequence reads with the same unique barcode ("*differentiator tag*") to determine the presence of "PCR errors," EX1012(McCloskey), [0063], and "mutations [in the DNA] arising during PCR amplification," *id.*, [0106]. McCloskey also explains that "[b]arcoding permits . . . correcting for skewing that can arise from PCR amplification or the cloning of products." *Id.*, [0005]. Thus, McCloskey teaches claim 3.

327. Bentley and McCloskey also teach claim 3. Using Bentley's technique, a skilled person would have grouped amplified nucleic acids with the same barcode and used them to determine a consensus sequence. Therefore, the combination of Bignell, McCloskey, and Bentley also render claim 3 obvious.

F. Bignell, McCloskey, and Gahan Render Obvious Claims 6 and 8

1. A Person of Ordinary Skill in the Art Would Have Practiced Bignell and McCloskey to Analyze Nucleic Acids in Blood For Diagnostic Purposes

328. McCloskey teaches that its method can be used to analyze "[h]uman genomic DNA [that] was isolated from blood." EX1012(McCloskey), [0045]. A person of ordinary skill would have known that different types of nucleic acids circulate in blood, and that it was routine to study these types of nucleic acids. It was also well-known that circulating nucleic acids from blood samples could be

analyzed for several diagnostic purposes, including for fetal medicine as well as the diagnosis of cancer and other diseases.

329. A person of ordinary skill would have known that it was useful to sequence nucleic acids circulating in blood for diagnostic purposes. That knowledge is reflected in Gahan, which explains that sequencing nucleic acids in blood can allow for early diagnosis of certain diseases. EX1015(Gahan), 1. Therefore, a person of ordinary skill would have been motivated to isolate the different types of nucleic acids circulating in blood and analyze them using Bignell's and McCloskey's method.

2. Claim 6

330. Claim 6 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules are circulating tumor nucleic acid molecules*."

331. McCloskey teaches that its method can be used to analyze "[h]uman genomic DNA [that] was isolated from blood." EX1012(McCloskey), [0045]. A person of ordinary skill in the art would have understood that a blood sample contains circulating nucleic acid molecules. A person of ordinary skill in the art would have also known that tumor cells shed DNA into the bloodstream. Gahan, for instance, explains that "[c]ells releasing these forms of DNAs could include tissue-based cells related to . . . tumors." EX1015(Gahan), 2.

332. A person of ordinary skill in the art would have understood that, if a tumor is present, the nucleic acid molecules in blood samples can comprise circulating tumor nucleic acid molecules. *See, e.g.*, EX1032, 654 ("detection of mutations in tumor cells . . . in the circulation"). Gahan reports that analyzing circulating nucleic acids can identify breast cancer and lung cancer tumors. EX1015(Gahan), 4. Thus, a person of ordinary skill in the art would have found it obvious to use McCloskey's methods to analyze "*circulating tumor nucleic acid molecule[s]*," for diagnostic purposes as described by Gahan.

3. Claim 8

333. Claim 8 depends on claim 1 and further recites "*wherein said plurality of target nucleic acid molecules are maternally-circulating fetal or placental nucleic acid molecules*."

334. McCloskey teaches that its method can be used to analyze "[h]uman genomic DNA [that] was isolated from blood." EX1012(McCloskey), [0045]. As I explained above, Gahan teaches that fetal DNA from maternal blood, specifically blood plasma, can be analyzed. For instance, Gahan reports that "[s]everal studies have shown that the sex of the fetus could be determined using fetal DNA in maternal plasma, many of which have used the presence or absence of Y chromosome markers." EX1015(Gahan), 3-4. This means that a maternal blood sample comprises circulating fetal cells which can be analyzed, such as to

determine the sex of the fetus. Gahan also teaches that maternal blood plasma can be analyzed to detect maternal complications, such as "fetal disorders." *Id.*, 3. Thus, a person of ordinary skill in the art would have found it obvious to use McCloskey's methods to analyze "*maternally-circulating fetal or placental nucleic acid molecules*," as described by Gahan.

G. Bignell, McCloskey, and Chiang Render Obvious Claim 7

1. Overview of Chiang

335. Chiang teaches a method of detecting copy number using Next-Gen sequencing techniques. EX1033(Chiang), 99 (discussing "High-resolution mapping of copy number alterations with massively parallel sequencing"). Chiang also explains that "copy number variants can be used for phenotypic mapping in genome-wide association studies and have been linked to various diseases," including cancer. *Id.* Chiang describes algorithms for determining copy number using using massively parallel sequencing, and it tested those algorithms on an Illumina Genome Analyzer sequencer. *Id.*, 100-101. The authors found their approach had over two times higher resolution than then-existing microarray techniques. *Id.*, 102.

336. Chiang analyzes sequence reads from human cell lines derived from tumor samples. EX1033(Chiang), 99, 100 ("tumor sample"), 101 ("tumor cell lines"). The sequence reads were generated using Illumina's Genome Analyzer.

Id., 99, 101. Chiang reports that "copy number estimates from massively parallel sequencing achieved greater sensitivity, higher dynamic range and greater precision for mapping breakpoints than similar estimates based on microarray hybridization." *Id.*, 99. Even so, Chiang recognizes that it "anticipate[s] that each massively parallel sequencing platform may be susceptible to particular biases." *Id.*, 102.

2. A Person of Ordinary Skill in the Art Would Have Practiced Bignell and McCloskey to Determine Copy Number Taught by Chiang

337. A person of ordinary skill in the art would have been motivated to detect copy number with Next-Gen sequencing techniques. Chiang teaches a method of detecting copy number using Next-Gen sequencing techniques. EX1033(Chiang), 99 (discussing "High-resolution mapping of copy number alterations with massively parallel sequencing"). A person of ordinary skill would have understood that determining copy number was an "important application" of next-generation sequencing. EX1014, 1143. Chiang explains that analyzing copy number "can be used for phenotypic mapping in genome-wide association studies and have been linked to numerous diseases" and that copy number is a "powerful way to find cancer genes." EX1033(Chiang), 99. Chiang also reports that "copy number estimates from massively parallel sequencing achieved greater sensitivity,

higher dynamic range and greater precision for mapping breakpoints than similar estimates based on microarray hybridization." *Id.*

338. A person of ordinary skill in the art would have also had a reasonable expectation of success in detecting copy number according to Chiang using the Next-Gen sequencing techniques taught by Bignell. Bignell teaches that sequencing can be performed using the "Illumina Genome Analyser." EX1013(Bignell), 45:48-50. Chiang's method uses sequencing data from Illumina's Genome Analyzer. EX1033(Chiang), 99.

339. A person of ordinary skill in the art would have also been motivated to combine McCloskey's method of correcting bias with Chiang. Chiang, recognizes that there are "issues involved in identifying cancer copy-number alterations using massively parallel sequencing." EX1033(Chiang), 99. One issue Chiang discusses is that the results may reflect "inherent biases in the sample preparation or sequencing procedures." *Id.*, 101; *see also id.*, 102 ("We anticipate that each massively parallel sequencing platform may be susceptible to particular biases"). McCloskey's methods provides a solution for the issue presented by Chiang. McCloskey explains that "[b]arcoding permits . . . correcting for skewing that can arise from PCR amplification or the cloning of products." EX1012(McCloskey), [0005]. Additionally, McCloskey teaches that its method "allows for the analysis of mutations arising during PCR amplification," *id.*,

[0106], as well as "PCR error," *id.*, [0063]. A skilled person would have also recognized that McCloskey's technique allowed for accurately determining the number of times a particular sequence appeared in the sample.

340. Therefore, a person of ordinary skill in the art would have been motivated to combine Chiang's method of detecting copy number with Bignell and McCloskey, and would have had a reasonable expectation of success in doing so.

3. Claim 7

341. Claim 7 depends on claim 1 and further recites "*further comprising the step of determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained*."

342. As I explained above, a person of skill in the art would have understood that it should rely on knowledge that was known in the art to determine "*the copy number of a genomic region*." *See* ¶¶ 209-210.

343. Chiang teaches a method for "*determining the copy number of a genomic region or transcript in a patient from whom the sample is obtained*." Chiang explains that the "*the copy number of a genomic region*" can be determined by calculating the "copy-number ratio." EX1033(Chiang), 100. The "copy-number ratio equals the number of aligning reads from a tumor sample, divided by the number from the corresponding matched normal sample." *Id.* When the "copy-number ratio deviates from 1," that is indicative of copy number variation in

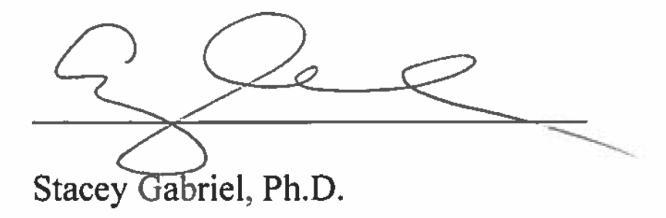
that region of the genome. EX1033(Chiang), 100. Chiang also teaches that it "developed a computational algorithm, called SeqSeq, to detect and localize copy number alterations from massively parallel sequence data." *Id.*

344. Chiang also teaches a method for detecting copy number using a sample from a patient. Chiang explains that it analyzed "a collection of ~14 million aligned sequence reads ***from human cell lines***." *Id.*, 99. Chiang's methods also analyze "tumor samples," EX1033(Chiang), 100, 102, and "three tumor cell lines (HCC1954, HCC1143 and NCI-H3247)," *id.*, 101. A person of ordinary skill in the art would have recognized that at least HCC1954 and HCC1143 were tumor cell lines derived from human patients.

* * *

I, Stacey Gabriel, do hereby declare and state, that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, under Section 1001 of Title 18 of the United States Code.

Executed on: September 30, 2024



Stacey Gabriel, Ph.D.

Exhibit List

The list of exhibits that I relied upon in forming my opinions is below.

Exhibit #	Reference Name
1001	U.S. 11,840,730 ("'730 patent)
1002	'862 Application FH
1003	'687 Application FH
1004	'764 Application FH
1006	Provisional Appl. No. 61/174,470
1007	Provisional Appl. No. 61/178,923
1008	Provisional Appl. No. 61/179,358
1009	Provisional Appl. No. 61/182,089
1010	US 9,085,798 to Chee
1011	Provisional Appl. No. 61/174,442
1012	US Pub. McCloskey, US2007/0020640 to McCloskey
1013	US 8,053,192 to Bignell
1014	J. Shendure & H. Ji, Next-Ggeneration DNA Ssequencing, Nature Biotechnology, Vol. 26, No. 10 (Oct. 2008)
1015	Gahan, Circulating Nucleic Acids in Plasma and Serum, Ann. N.Y. Acad. Sci. (2008)
1016	Weiner, Gabriel, Stephens, Eds. (2007), Genetic Variation: A Laboratory Manual (excerpts)
1017	D. Bentley, Whole-Genome Re-Sequencing, Current Opinion in Genetics & Development 2006, 16:545-552 (2006)
1018	US Pub. 2008/0261204 to Lexow
1019	US 6,172,214 to Brenner
1020	Brenner et al., *In vitro* Cloning of Complex Mixtures of DNA on Microbeads: Physical Separation of Differentially Expressed cDNAs, PNAS, Vol. 97, No. 4 (Feb. 2000)

Exhibit #	Reference Name
1025	S. Arcinas et al., PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample, APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2005, p. 8966-69, Vol. 71, No. 12
1029	Ronaghi, US Pub. 2006/0008824
1032	Voelkerding et al., Next-Generation Sequencing: From Basic Research to Diagnostics, Clinical Chemistry, Vol 55, No. 4 (April 2009)
1033	D. Chiang et al., High-resolution mapping of copy-number alterations with massively parallel sequencing, Nature Methods Vol. 6 No. 1 (Jan. 2009)
1034	Print out of Polymerase Chain Reaction (PCR), Genome.gov, available at: https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction
1035	Commins et al., Computational Biology Methods and Their Application to the Comparative Genomics of Endocellular Symbiotic Bacteria of Insects, Biological Procedures Online, Vol. 11, No. 1 (March 2009)

EXHIBIT A

Stacey Bolk Gabriel, Ph.D.
Legal name: Stacey Lynn Munsey
15 Manhattan Dr., Burlington, MA 01803
781- 254 -9469 (c) ~ Stacey@broadinstitute.org

Research and Leadership Experience

Broad Institute of MIT and Harvard (formerly Whitehead Institute Center for Genome Research), 11/98 - Present

Executive Vice President, Head of Platforms and Scientific Execution, Broad Institute of Harvard and MIT, 10/2024 - present

- Oversee Broad Institute platform organization, inclusive of nine units (Genomics, Data Science, Genomic Perturbation, Proteomics, Metabolomics, Development of Therapeutics, Spatial Technologies, Imagining and Patient Engagement) totaling ~500 staff.
- Set strategy, performance metrics, financial targets long term sustainability.
- Continue activities of Chief Genomics Officer (previous role) including serving as Principal Investigator on flagship programs.
- Member of the Broad Institute Executive Leadership Team (C-suite membership)

Chief Genomics Officer, Broad Institute and Senior Director, Genomics Platform, Broad Institute of Harvard and MIT, 5/2012 – 10/2024

- Direct the Broad Institute's largest platform, and the largest US genome center, comprising >200 people dedicated to all sample handling, microarray, genotyping, and sequencing and data management activities.
- Responsible for $150M (average) annual budget for genomic activities, expanded to >$600M annually with COVID-19 diagnostics.
- Oversee all strategy, investment, execution of genomics-related activities.
- Oversee technology development and evaluation and implementation of new technology platforms.
- Serve as Principal Investigator of flagship NIH program awards for All of Us Research Program (>$230M cumulative), Gabriella Kids First Research Program (>60M cumulative), NHLBI Trans-omics for Precision Medicine (>150M cumulative).
- Member of the Broad Institute Executive Leadership Team, serving as Chair of Professional Scientists.

Co-Director, Genome Sequence and Analysis Program and Medical and Population Genetics Program, Broad Institute of Harvard and MIT, 1/2009 – 5/2012

- Responsible for planning, execution and delivery of a portfolio of cancer and medical sequencing projects as part of the NHGRI Genome Sequencing Program with a Center budget total of ~$40M annually.
- As Principal Investigator (PI), secured over $250M in other NIH awards aimed at large scale genotyping and sequencing.

- Direct the activity of cross-disciplinary teams, totaling 60 people, including project managers, analysts, computational biologists and software engineers in the analysis of next-generation sequence data as applied to an array of cancer genomics and medical genetics projects.
- Served as Co-chair of Data Production committee for the International 1000 Genome Project, as well as serving as a member of the Executive and Steering committee for The Cancer Genome Atlas.

Director, **Genetic Analysis Platform,** Broad Institute of Harvard and MIT, 5/2004 – 1/2009
- Responsible for creating, scaling and directing the Genetic Analysis Platform of the Broad Institute. The platform encompasses all production and data management activities related to nucleic acid analysis (gene expression, genotyping, re-sequencing).
- At peak period (2006 – 2008) operated the platform with yearly revenue of $45 million, overseeing a staff of 65 individuals (project managers, research scientists, software engineers and computational biologists).
- Key milestones included producing microarray data on over 1000,000 DNA samples over an 18-month period.
- Directed data production for over 50 publications describing Genome-wide association findings.

Associate Director, High Throughput Biology, Medical and Population Genetics Program, Whitehead Institute Center for Genome Research, 5/2003 – 5/2004
- In this role spearheaded the expansion of SNP genotyping activity from targeted activity for Human HapMap project to a centralized technology platform with dedicated activity in technology development, large-scale production, data management and analysis.
- Oversaw successful completion of the Whitehead's contribution to the Human HapMap Project ($10 M project).

Scientific Director, SNP genotyping / HapMap Program, Whitehead Institute Center for Genome Research, 2/2002 – 5/2003
- Responsible for all aspects of the center's contribution to the International HapMap Project.
- Oversaw a team of 15 technicians, analysts and software engineers.
- Played active role in project design and quality control.
- Served on International HapMap Project Steering committee.

Research Scientist, Program in Medical and Population Genetics, Whitehead Institute Center for Genome Research, 11/98 – 2/2002
- Investigated the haplotype structure of the human genome. Research resulting in first author Science publication, widely regarded as laying the foundation for the Human HapMap Project.
- Technical development and implementation of the first platforms to be used for large high throughput SNP genotyping

Graduate Student, Department of Genetics, Case Western Reserve University, Cleveland, OH, 1/94– 11/98 Advisor: Aravinda Chakravarti, Ph.D.

- Dissertation: "Genetic Dissection of Complex Human Disease"

Research Assistant, Department of Human Genetics, Chakravarti Lab, University of Pittsburgh, Pittsburgh, PA, 5/92 – 1/94

- Conducted lab work and phlebotomy in support of human genetic studies

Education and degrees

Doctor of Philosophy, Genetics. Case Western Reserve University, Department of Genetics, 1998.

Dissertation: "Hirschsprung disease: Genetic dissection of a complex phenotype." Advisor: Aravinda Chakravarti.

Bachelor of Science, Molecular Biology. Carnegie Mellon University, Pittsburgh, PA, 1993.

Peer Review and other Extramural Activities

February 2007 – December 2010	External Advisory Committee, NHLBI Resequencing and Genotyping Service
November 2008 – 2009	Editorial Board, Human Genetics
March 2008 – 2012	External Advisory Committee, USC Epigenome Center
February 2009	NIH Genomics, Computational Biology and Technology Study Section, Ad Hoc reviewer
March 2009	Editorial Board, Genome Research
July 2009	NHLBI Special Emphasis Panel
July, 2009 – 2013	Standing Member, NIH Genomics, Computational Biology and Technology Study Section
January 2010 – 2016	Member, Genome Canada Science and Industry Advisory Committee
August 2011 – 2016	Member, Scientific Advisory Board, PAAR4KIDS, St. Jude Research Hospital
August 2012 – 2016	Consultant, KEW Group
June 2015 – 2017	Member, Scientific Advisory Board, RainDance Technologies
June 2016 – December 2020	Expert Witness, Foundation Medicine
July 2013 – Present	Expert Witness, Perkins Coie
June 2019 – Present	Chair, Research Oversight Committee, PERSPECTIVE Study, Genome Canada
Sept 2022 – Present	Member, Scientific Advisory Board, Frederick National Lab

May 2023 – October 2024	Expert Witness, WilmerHale
July 2024 – Present	Expert Witness, Sidley Austin LLP

Awards and Recognition

Most Influential Scientific Minds*, Thomas Reuters (2014, 2015)
*Highest ranking of researchers across disciplines

Highly Cited Researcher, Clarivate (2014 - 2021)

American Association of Cancer Research, Team Science Award (2020)

Bostonians of the Year, Boston Globe Magazine (2020)

ISI Highly Cited Researcher (Current)

Honorary Doctor of Science, Tufts University (2021)

Distinguished Alumni Award in Medicine, Case Western Reserve University School of Medicine (2021)

Medal of Distinction, Barnard College, (2022)

Investigator of the Year, Massachusetts Society for Medical Research (2022)

Best Female Scientists in the World 2022, Ranking #22, Research.com (2022)

Publications (277 Total)

1. Jiang MZ, Aguet F, Ardlie K, Chen J, Cornell E, Cruz D, Durda P, Gabriel SB, Gerszten RE, et al. Canonical correlation analysis for multi-omics: Application to cross-cohort analysis. PLoS Genet. 2023 May 22;19(5):e1010517.

2. Seyerle AA, Laurie CA, Coombes BJ, Jain D, Conomos MP, Brody J, Chen MH, Gogarten SM, Beutel KM, et al. Whole Genome Analysis of Venous Thromboembolism: the Trans-Omics for Precision Medicine Program. Circ Genom Precis Med. 2023 Apr;16(2):e003532.

3. Sazonovs A, Stevens CR, Venkataraman GR, Yuan K, Avila B, Abreu MT, Ahmad T, Allez M, Ananthakrishnan AN, Atzmon G, Baras A, Barrett JC, Barzilai N,et al. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. Nat Genet. 2022 Sep;54(9):1275-1283.

4. Siddle KJ, Krasilnikova LA, Moreno GK, Schaffner SF, Vostok J, Fitzgerald NA,

Lemieux JE, Barkas N, Loreth C, Specht I, Tomkins-Tinch CH, Paull JS, Schaeffer B, Taylor BP, Loftness B, Johnson H, Schubert PL, Shephard HM, Doucette M, Fink T, Lang AS, Baez S, Beauchamp J, Hennigan S, Buzby E, Ash S, Brown J, Clancy S, Cofsky S, Gagne L, Hall J, Harrington R, Gionet GL, DeRuff KC, Vodzak ME, Adams GC, Dobbins ST, Slack SD, Reilly SK, Anderson LM, Cipicchio MC, DeFelice MT, Grimsby JL, Anderson SE, Blumenstiel BS, Meldrim JC, Rooke HM, Vicente G, Smith NL, Messer KS, Reagan FL, Mandese ZM, Lee MD, Ray MC, Fisher ME, Ulcena MA, Nolet CM, English SE, Larkin KL, Vernest K, Chaluvadi S, Arvidson D, Melchiono M, Covell T, Harik V, Brock-Fisher T, Dunn M, Kearns A, Hanage WP, Bernard C, Philippakis A, Lennon NJ, Gabriel SB, Gallagher GR, Smole S, Madoff LC, Brown CM, Park DJ, MacInnis BL, Sabeti PC. Transmission from vaccinated individuals in a large SARS-CoV-2 Delta variant outbreak. Cell. 2022 Feb 3;185(3):485-492.

5. Simas AM, Crott JW, Sedore C, Rohrbach A, Monaco AP, Gabriel SB, Lennon N, Blumenstiel B, Genco CA. Pooling for SARS-CoV2 Surveillance: Validation and Strategy for Implementation in K-12 Schools. Front Public Health. 2021 Dec 17;9:789402. doi: 10.3389/fpubh.2021.789402. PMID: 34976934; PMCID: PMC8718607.

6. Honigberg MC, Zekavat SM, Niroula A, Griffin GK, Bick AG, Pirruccello JP, *et al.* Premature Menopause, Clonal Hematopoiesis, and Coronary Artery Disease in Postmenopausal Women. *Circulation* 2021;**143**:410–23.

7. Zeng Z, Fu J, Cibulskis C, Jhaveri A, Gumbs C, Das B, *et al.* Cross-Site Concordance Evaluation of Tumor DNA and RNA Sequencing Platforms for the CIMAC-CIDC Network. *Clin Cancer Res* 2021;**27**:5049–61.

8. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021;**590**:290–9.

9. Cleary B, Hay JA, Blumenstiel B, Harden M, Cipicchio M, Bezney J, *et al.* Using viral load and epidemic dynamics to optimize pooled testing in resource-constrained settings. *Sci Transl Med* 2021;**13.**: https://doi.org/10.1126/scitranslmed.abf1568.

10. Pollock NR, Jacobs JR, Tran K, Cranston AE, Smith S, O'Kane CY, *et al.* Performance and Implementation Evaluation of the Abbott BinaxNOW Rapid Antigen Test in a High-Throughput Drive-Through Community Testing Site in Massachusetts. *J Clin Microbiol* 2021;**59.**: https://doi.org/10.1128/JCM.00083-21.
11. Penter L, Zhang Y, Savell A, Huang T, Cieri N, Thrash EM, *et al.* Molecular and cellular features of CTLA-4 blockade for relapsed myeloid malignancies after transplantation. *Blood* 2021;**137**:3212–7.

12. Natarajan P, Pampana A, Graham SE, Ruotsalainen SE, Perry JA, de Vries PS, *et al.* Chromosome Xq23 is associated with lower atherogenic lipid concentrations and

favorable cardiometabolic indices. *Nat Commun* 2021;**12**:2182.

13. Hu Y, Stilp AM, McHugh CP, Rao S, Jain D, Zheng X, *et al.* Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *Am J Hum Genet* 2021;**108**:874–93.

14. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, *et al.* Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 2021;**373.**: https://doi.org/10.1126/science.abh0635.

15. Hu Y, Stilp AM, McHugh CP, Rao S, Jain D, Zheng X, *et al.* Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *Am J Hum Genet* 2021;**108**:1165.

16. Pollock NR, Berlin D, Smole SC, Madoff LC, Brown C, Henderson K, *et al.* Implementation of SARS-CoV2 Screening in K-12 Schools Using In-School Pooled Molecular Testing and Deconvolution by Rapid Antigen Test. *J Clin Microbiol* 2021;**59**:e0112321.

17. Pollock NR, Tran K, Jacobs JR, Cranston AE, Smith S, O'Kane CY, *et al.* Performance and Operational Evaluation of the Access Bio CareStart Rapid Antigen Test in a High-Throughput Drive-Through Community Testing Site in Massachusetts. *Open Forum Infect Dis* 2021;**8**:ofab243.

18. Lennon NJ, Bhattacharyya RP, Mina MJ, Rehm HL, Hung DT, Smole S, *et al.* Cross-sectional assessment of SARS-CoV-2 viral load by symptom status in Massachusetts congregate living facilities. *J Infect Dis* 2021. https://doi.org/10.1093/infdis/jiab367.

19. Brown CM, Vostok J, Johnson H, Burns M, Gharpure R, Sami S, *et al.* Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings - Barnstable County, Massachusetts, July 2021. *MMWR Morb Mortal Wkly Rep* 2021;**70**:1059–62.
20. Bick AG, Pirruccello JP, Griffin GK, Gupta N, Gabriel S, Saleheen D, *et al.* Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* 2020;**141**:124–31.

21. McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, Furlong EEM, *et al.* The road ahead in genetics and genomics. *Nat Rev Genet* 2020;**21**:581–96.

22. Nomura A, Emdin CA, Won HH, Peloso GM, Natarajan P, Ardissino D, *et al.* Heterozygous ABCG5 Gene Deficiency and Risk of Coronary Artery Disease. *Circ Genom Precis Med* 2020;**13**:417–23.

23. Zhao X, Qiao D, Yang C, Kasela S, Kim W, Ma Y, *et al.* Whole genome sequence analysis of pulmonary function and COPD in 19,996 multi-ethnic participants. *Nat Commun* 2020;**11**:5182.

24. Emdin CA, Haas ME, Khera AV, Aragam K, Chaffin M, Klarin D, *et al.* A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. *PLoS Genet* 2020;**16**:e1008629.

25. Verboon JM, Mahmut D, Kim AR, Nakamura M, Abdulhay NJ, Nandakumar SK, *et al.* Infantile Myelofibrosis and Myeloproliferation with CDC42 Dysfunction. *J Clin Immunol* 2020. https://doi.org/10.1007/s10875-020-00778-7.

26. Wang Q, Pierce-Hoffman E, Cummings BB, Alföldi J, Francioli LC, Gauthier LD, *et al.* Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nat Commun* 2020;**11**:2539.

27. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, *et al.* Characterizing the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nat Commun* 2020;**11**:2523.

28. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, *et al.* A structural variation reference for medical and population genetics. *Nature* 2020;**581**:444–51.

29. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, *et al.* Evaluating drug targets through human loss-of-function genetic variation. *Nature* 2020;**581**:459–64.

30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43.

31. Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, *et al.* Transcript expression-aware annotation improves rare variant interpretation. *Nature* 2020;**581**:452–8.

32. Whiffin N, Armean IM, Kleinman A, Marshall JL, Minikel EV, Goodrich JK, *et al.* The effect of LRRK2 loss-of-function variants in humans. *Nat Med* 2020;**26**:869–77.

33. Cleary B, Hay JA, Blumenstiel B, Gabriel S, Regev A, Mina MJ. Efficient prevalence estimation and infected sample identification with group testing for SARS-CoV-2. *medRxiv* 2020. https://doi.org/10.1101/2020.05.01.20086801.

34. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019;**570**:71–6.

35. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* 2019;**565**:234–9.

36. Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, *et al.* Whole-Genome Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction. *Circulation* 2019;**139**:1593–602.

37. Abdulhay NJ, Fiorini C, Verboon JM, Ludwig LS, Ulirsch JC, Zieger B, *et al.* Impaired human hematopoiesis due to a cryptic intronic GATA1 splicing mutation. *J Exp Med* 2019;**216**:1050–60.

38. Tilghman JM, Ling AY, Turner TN, Sosa MX, Krumm N, Chatterjee S, *et al.* Molecular Genetic Anatomy and Risk Profile of Hirschsprung's Disease. *N Engl J Med* 2019;**380**:1421–32.

39. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019;**570**:71–6.

40. Bick AG, Pirruccello JP, Griffin GK, Gupta N, Gabriel S, Saleheen D, *et al.* Genetic IL-6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* 2019. https://doi.org/10.1161/CIRCULATIONAHA.119.044362.

41. Chatterjee S, Nandakumar P, Auer DR, Gabriel SB, Chakravarti A. Gene- and tissue-level interactions in normal gastrointestinal development and Hirschsprung disease. *Proc Natl Acad Sci U S A* 2019. https://doi.org/10.1073/pnas.1908756116.

42. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. PLoS Genet 2019;**15**:e1008500.

43. Alver M, Palover M, Saar A, Läll K, Zekavat SM, Tõnisson N, *et al.* Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. *Genet Med* 2019;**21**:1173–80.

44. Emdin CA, Khera AV, Aragam K, Haas M, Chaffin M, Klarin D, *et al.* DNA Sequence Variation in ACVR1C Encoding the Activin Receptor-Like Kinase 7 Influences Body Fat Distribution and Protects Against Type 2 Diabetes. *Diabetes* 2019;**68**:226–34.

45. Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, *et al.* Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun* 2018;**9**:1613.

46. Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, *et al.* Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 2018;**19**:332.

47. Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, *et al.* Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell* 2018;**174**:433–47.e19.

48. Bis JC, Jian X, Kunkle BW, Chen Y, Hamilton-Nelson KL, Bush WS, *et al.* Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol Psychiatry* 2018. https://doi.org/10.1038/s41380-018-0112-7.

49. Guemez-Gamboa A, Çağlayan AO, Stanley V, Gregor A, Zaki MS, Saleem SN, *et al.* Loss of Protocadherin-12 Leads to Diencephalic-Mesencephalic Junction Dysplasia Syndrome. *Ann Neurol* 2018;**84**:638–47.

50. Khalil R, Kenny C, Hill RS, Mochida GH, Nasir R, Partlow JN, *et al.* PSMD12 haploinsufficiency in a neurodevelopmental disorder with autistic features. *Am J Med Genet B Neuropsychiatr Genet* 2018;**177**:736–45.

51. Ulirsch JC, Verboon JM, Kazerounian S, Guo MH, Yuan D, Ludwig LS, *et al.* The Genetic Landscape of Diamond-Blackfan Anemia. *Am J Hum Genet* 2018;**103**:930–47.

52. Choi SH, Weng L-C, Roselli C, Lin H, Haggerty CM, Shoemaker MB, *et al.* Association Between Titin Loss-of-Function Variants and Early-Onset Atrial Fibrillation. *JAMA* 2018;**320**:2354–64.

53. Guo MH, Nandakumar SK, Ulirsch JC, Zekavat SM, Buenrostro JD, Natarajan P, *et al.* Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. *Proc Natl Acad Sci U S A* 2017;**114**:E327–36.

54. Lardelli RM, Schaffer AE, Eggens VRC, Zaki MS, Grainger S, Sathe S, *et al.* Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat Genet* 2017;**49**:457–64.

55. Ghazani AA, Oliver NM, St Pierre JP, Garofalo A, Rainville IR, Hiller E, *et al.* Assigning clinical meaning to somatic and germ-line whole-exome sequencing data in a prospective cancer precision medicine study. *Genet Med* 2017;**19**:787–95.

56. Khera AV, Won H-H, Peloso GM, O'Dushlaine C, Liu D, Stitziel NO, *et al.* Association of Rare and Common Variation in the Lipoprotein Lipase Gene With Coronary Artery Disease. *JAMA* 2017;**317**:937–46.

57. Kim AR, Ulirsch JC, Wilmes S, Unal E, Moraga I, Karakukcu M, *et al.* Functional Selectivity in Cytokine Signaling Revealed Through a Pathogenic EPO Mutation. *Cell* 2017;**168**:1053–64.e15.

58. Manning A, Highland HM, Gasser J, Sim X, Tukiainen T, Fontanillas P, *et al.* A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. *Diabetes* 2017;**66**:2019–32.

59. Stitziel NO, Khera AV, Wang X, Bierhals AJ, Vourakis AC, Sperry AE, *et al.* ANGPTL3 Deficiency and Protection Against Coronary Artery Disease. *J Am Coll Cardiol* 2017;**69**:2054–63.

60. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet* 2017;**25**:869–76.

61. Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 2017;**544**:235–9.

62. Nomura A, Won H-H, Khera AV, Takeuchi F, Ito K, McCarthy S, *et al.* Protein-Truncating Variants at the Cholesteryl Ester Transfer Protein Gene and Risk for Coronary Heart Disease. *Circ Res* 2017;**121**:81–8.

63. Huang FW, Mosquera JM, Garofalo A, Oh C, Baco M, Amin-Mansour A, *et al.* Exome Sequencing of African-American Prostate Cancer Reveals Loss-of-Function ERF Mutations. *Cancer Discov* 2017;**7**:973–83.

64. Johnson CP, Kim IK, Esmaeli B, Amin-Mansour A, Treacy DJ, Carter SL, *et al.* Systematic genomic and translational efficiency studies of uveal melanoma. *PLoS One* 2017;**12**:e0178189.

65. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* 2017;**377**:111–21.

66. Rheinbay E, Parasuraman P, Grimsby J, Tiao G, Engreitz JM, Kim J, *et al.* Recurrent and functional regulatory mutations in breast cancer. *Nature* 2017;**547**:55–60.

67. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017;**547**:217–21.

68. Mercader JM, Liao RG, Bell AD, Dymek Z, Estrada K, Tukiainen T, *et al.* A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for Type 2 Diabetes. *Diabetes* 2017;**66**:2903–14.

69. Emdin CA, Khera AV, Klarin D, Natarajan P, Zekavat SM, Nomura A, *et al.* Phenotypic Consequences of a Genetic Predisposition to Enhanced Nitric Oxide Signaling. *Circulation* 2018;**137**:222–32.

70. Brody JA, Morrison AC, Bis JC, O'Connell JR, Brown MR, Huffman JE, *et al.* Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet* 2017;**49**:1560–3.

71. Flannick J, Fuchsberger C, Mahajan A, Teslovich TM, Agarwala V, Gaulton KJ, *et al.* Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. *Sci Data* 2017;**4**:170179.

72. Cancer Genome Atlas Research Network, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, *et al.* Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* 2016;**374**:135–45.

73. Kan M, Auer PL, Wang GT, Bucasas KL, Hooker S, Rodriguez A, *et al.* Rare variant associations with waist-to-hip ratio in European-American and African-American women from the NHLBI-Exome Sequencing Project. *Eur J Hum Genet* 2016;**24**:1181–7.

74. Khera AV, Won H-H, Peloso GM, Lawson KS, Bartz TM, Deng X, *et al.* Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *J Am Coll Cardiol* 2016;**67**:2578–89.

75. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, *et al.* Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* 2016;**15**:857–65.

76. Blumenstiel B, DeFelice M, Birsoy O, Bleyer AJ, Kmoch S, Carter TA, *et al.* Development and Validation of a Mass Spectrometry-Based Assay for the Molecular Diagnosis of Mucin-1 Kidney Disease. *J Mol Diagn* 2016;**18**:566–71.

77. Burger JA, Landau DA, Taylor-Weiner A, Bozic I, Zhang H, Sarosiek K, *et al.* Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat Commun* 2016;**7**:11589.

78. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, *et al.* The genetic architecture of type 2 diabetes. *Nature* 2016;**536**:41–7.

79. Peloso GM, Lange LA, Varga TV, Nickerson DA, Smith JD, Griswold ME, *et al.* Association of Exome Sequences With Cardiovascular Traits Among Blacks in the Jackson Heart Study. *Circ Cardiovasc Genet* 2016;**9**:368–74.

80. Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, *et al.* Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet* 2016;**48**:1071–6.

81. Li H, Bielas SL, Zaki MS, Ismail S, Farfara D, Um K, *et al.* Biallelic Mutations in Citron Kinase Link Mitotic Cytokinesis to Human Primary Microcephaly. *Am J Hum Genet* 2016;**99**:501–10.

82. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–91.

83. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;**48**:1279–83.

84. Chatterjee S, Kapoor A, Akiyama JA, Auer DR, Lee D, Gabriel S, *et al.* Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. *Cell* 2016;**167**:355–68.e10.

85. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, *et al.* Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* 2016;**17**:1206.

86. Lennon NJ, Adalsteinsson VA, Gabriel SB. Technological considerations for genome-guided diagnosis and management of cancer. *Genome Med* 2016;**8**:112.

87. Taylor-Weiner A, Zack T, O'Donnell E, Guerriero JL, Bernard B, Reddy A, *et al.* Genomic evolution and chemoresistance in germ-cell tumours. *Nature* 2016;**540**:114–8.

88. Emdin CA, Khera AV, Natarajan P, Klarin D, Won H-H, Peloso GM, *et al.* Phenotypic Characterization of Genetically Lowered Human Lipoprotein(a) Levels. *J Am Coll Cardiol* 2016;**68**:2761–72.

89. 177 Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ, Neale BM. Phenotypic extremes in rare variant study designs. *Eur J Hum Genet* 2016;**24**:924–30.

90. Schick UM, Auer PL, Bis JC, Lin H, Wei P, Pankratz N, *et al.* Association of exome sequences with plasma C-reactive protein levels in >9000 participants. *Hum Mol Genet* 2015;**24**:559–71.

91. Do R, Stitziel NO, Won H-H, Jørgensen AB, Duga S, Angelica Merlini P, *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* 2015;**518**:102–6.

92. 162 Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, *et al.* Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science* 2015;**347**:1258522.

93. 163 Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, *et al.* Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med* 2015;**7**:270ra6.
94. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, *et al.* Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet* 2015;**11**:e1004876.

95. Stitziel NO, Peloso GM, Abifadel M, Cefalu AB, Fouchier S, Motazacker MM, *et al.* Exome sequencing in suspected monogenic dyslipidemias. *Circ Cardiovasc Genet* 2015;**8**:343–50.

96. Thormaehlen AS, Schuberth C, Won H-H, Blattmann P, Joggerst-Thomalla B, Theiss S, *et al.* Systematic cell-based phenotyping of missense alleles empowers rare variant association studies: a case for LDLR and myocardial infarction. *PLoS Genet* 2015;**11**:e1004855.

97. Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res* 2015;**25**:316–27.

98. Yuen M, Sandaradura SA, Dowling JJ, Kostyukova AS, Moroz N, Quinlan KG, *et al.* Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J Clin Invest* 2015;**125**:456–7.

99. Wilson FH, Johannessen CM, Piccioni F, Tamayo P, Kim JW, Van Allen EM, *et al.* A functional landscape of resistance to ALK inhibition in lung cancer. *Cancer Cell* 2015;**27**:397–408.

100. Akizu N, Cantagrel V, Zaki MS, Al-Gazali L, Wang X, Rosti RO, *et al.* Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-autophagosome dysfunction. *Nat Genet* 2015;**47**:528–34.

101. Guemez-Gamboa A, Nguyen LN, Yang H, Zaki MS, Kara M, Ben-Omran T, *et al.* Inactivating mutations in MFSD2A, required for omega-3 fatty acid transport in brain, cause a lethal microcephaly syndrome. *Nat Genet* 2015;**47**:809–13.

102. Roosing S, Hofree M, Kim S, Scott E, Copeland B, Romani M, *et al.* Functional genome-wide siRNA screen identifies KIAA0586 as mutated in Joubert syndrome. *Elife* 2015;**4**:e06602.

103. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RGW, Aldape KD, Yung WKA, Salama SR, *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* 2015;**372**:2481–98.

104. Rickman KA, Lach FP, Abhyankar A, Donovan FX, Sanborn EM, Kennedy JA, *et al.* Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *Cell Rep* 2015;**12**:35–41.

105. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet* 2015;**47**:1047–55.

106. Wang AT, Kim T, Wagner JE, Conti BA, Lach FP, Huang AL, *et al.* A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol Cell* 2015;**59**:478–90.

107. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 2015;**350**:207–11.

108. Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* 2015;**5**:1164–77.

109. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* 2015;**526**:525–30.

110. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, *et al.* Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 2014;**26**:813–25.

111. Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, *et al.* The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* 2014;**4**:94–109.

112. Wagle N, Van Allen EM, Treacy DJ, Frederick DT, Cooper ZA, Taylor-Weiner A, *et al.* MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. *Cancer Discov* 2014;**4**:61–8.

113. Akizu N, Silhavy JL, Rosti RO, Scott E, Fenstermaker AG, Schroth J, *et al.* Mutations in CSPP1 lead to classical Joubert syndrome. *Am J Hum Genet* 2014;**94**:80–6.

114. Ito K, Bick AG, Flannick J, Friedman DJ, Genovese G, Parfenov MG, *et al.* Increased burden of cardiovascular disease in carriers of APOL1 genetic variants. *Circ Res* 2014;**114**:845–50.

115. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, *et al.* Landscape of genomic alterations in cervical carcinomas. *Nature* 2014;**506**:371–5.

116. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;**505**:495–501.

117. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014;**506**:185–90.

118. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang Z-Z, *et al.* Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 2014;**94**:233–45.

119. Flannick J, Thorleifsson G, Beer NL, Jacobs SBR, Grarup N, Burtt NP, *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 2014;**46**:357–63.

120. Wagle N, Grabiner BC, Van Allen EM, Hodis E, Jacobus S, Supko JG, *et al.* Activating mTOR mutations in a patient with an extraordinary response on a phase I trial of everolimus and pazopanib. *Cancer Discov* 2014;**4**:546–53.

121. McFadden DG, Papagiannakopoulos T, Taylor-Weiner A, Stewart C, Carter SL, Cibulskis K, *et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 2014;**156**:1298–311.

122. Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, *et al.* Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. *Hum Genet* 2014;**133**:985–95.

123. Schaffer AE, Eggens VRC, Caglayan AO, Reuter MS, Scott E, Coufal NG, *et al.* CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell* 2014;**157**:651–63.

124. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med* 2014;**20**:682–8.

125. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, *et al.* Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 2014;**124**:453–62.

126. Cho J, Bass AJ, Lawrence MS, Cibulskis K, Cho A, Lee S-N, *et al.* Colon cancer-derived oncogenic EGFR G724S mutant identified by whole genome sequence analysis is dependent on asymmetric dimerization and sensitive to cetuximab. *Mol Cancer* 2014;**13**:141.

127. SIGMA Type 2 Diabetes Consortium, Estrada K, Aukrust I, Bjørkhaug L, Burtt NP, Mercader JM, *et al.* Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA* 2014;**311**:2305–14.

128. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, *et al.* Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 2014;**371**:22–31.

129. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* 2014;**10**:e1004494.

130. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014;**46**:944–50.

131. Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* 2014;**4**:1140–53.

132. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* 2014;**17**:1156–63.

133. Majithia AR, Flannick J, Shahinian P, Guo M, Bray M-A, Fontanillas P, *et al.* Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A* 2014;**111**:13127–32.

134. Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, *et al.* RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet* 2014;**46**:1264–6.

135. Myocardial Infarction Genetics Consortium Investigators, Stitziel NO, Won H-H, Morrison AC, Peloso GM, Do R, *et al.* Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* 2014;**371**:2072–82.

136. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014;**371**:2488–98.

137. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014;**371**:2477–87.

138. Law R, Dixon-Salazar T, Jerber J, Cai N, Abbasi AA, Zaki MS, *et al.* Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein Formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *Am J Hum Genet* 2014;**95**:721–8.

139. Perry JA, Kiezun A, Tonzi P, Van Allen EM, Carter SL, Baca SC, *et al.* Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc Natl Acad Sci U S A* 2014;**111**:E5564–73.

140. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;**43**:11.10.1–33.

141. Drier Y, Lawrence MS, Carter SL, Stewart C, Gabriel SB, Lander ES, *et al.* Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res* 2013;**23**:228–35.

142. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013;**493**:216–20.

143. Verhaak RGW, Tamayo P, Yang J-Y, Hubbard D, Zhang H, Creighton CJ, *et al.* Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest* 2013;**123**:517–25.

144. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* 2013;**41**:e67.

145. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* 2013;**45**:279–84.

146. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, *et al.* Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 2013;**77**:235–42.

147. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, *et al.* Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 2013;**77**:259–73.

148. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;**31**:213–9.

149. Kirby A, Gnirke A, Jaffe DB, Barešová V, Pochet N, Blumenstiel B, *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* 2013;**45**:299–303.

150. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 2013;**152**:714–26.

151. Akizu N, Shembesh NM, Ben-Omran T, Bastaki L, Al-Tawari A, Zaki MS, *et al.* Whole-exome sequencing identifies mutated c12orf57 in recessive corpus callosum hypoplasia. *Am J Hum Genet* 2013;**92**:392–400.

152. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* 2013;**45**:478–86.

153. Liu L, Sabo A, Neale BM, Nagaswamy U, Stevens C, Lim E, *et al.* Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet* 2013;**9**:e1003443.

154. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, *et al.* Punctuated evolution of prostate cancer genomes. *Cell* 2013;**153**:666–77.

155. Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013;**368**:2059–74.

156. Cefalù AB, Pirruccello JP, Noto D, Gabriel S, Valenti V, Gupta N, *et al.* A novel APOB mutation identified by exome sequencing cosegregates with steatosis, liver cancer, and hypocholesterolemia. *Arterioscler Thromb Vasc Biol* 2013;**33**:2021–5.

157. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;**499**:214–8.

158. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;**45**:1134–40.

159. Stitziel NO, Fouchier SW, Sjouke B, Peloso GM, Moscoso AM, Auer PL, *et al.* Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. *Arterioscler Thromb Vasc Biol* 2013;**33**:2909–14.

160. Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, *et al.* Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat Genet* 2013;**45**:1380–5.

161. Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, *et al.* The somatic genomic landscape of glioblastoma. *Cell* 2013;**155**:462–77.

162. Mannstadt M, Harris M, Bravenboer B, Chitturi S, Dreijerink KMA, Lambright DG, *et al.* Germline mutations affecting Gα11 in hypoparathyroidism. *N Engl J Med* 2013;**368**:2532–4.

163. Radmanesh F, Caglayan AO, Silhavy JL, Yilmaz C, Cantagrel V, Omar T, *et al.* Mutations in LAMB1 cause cobblestone brain malformation without muscular or ocular abnormalities. *Am J Hum Genet* 2013;**92**:468–74.

164. Travaglini L, Brancati F, Silhavy J, Iannicelli M, Nickerson E, Elkhartoufi N, *et al.* Phenotypic spectrum and prevalence of INPP5E mutations in Joubert syndrome and related disorders. *Eur J Hum Genet* 2013;**21**:1074–8.

165. Hoshida Y, Villanueva A, Sangiovanni A, Sole M, Hur C, Andersson KL, *et al.* Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. *Gastroenterology* 2013;**144**:1024–30.

166. Puffenberger EG, Jinks RN, Sougnez C, Cibulskis K, Willert RA, Achilly NP, *et al.* Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One* 2012;**7**:e28936.

167. Lee JH, Silhavy JL, Lee JE, Al-Gazali L, Thomas S, Davis EE, *et al.* Evolutionarily assembled cis-regulatory module at a human ciliopathy locus. *Science* 2012;**335**:966–9.

168. Motazacker MM, Pirruccello J, Huijgen R, Do R, Gabriel S, Peter J, *et al.* Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia. *Eur Heart J* 2012;**33**:1360–6.

169. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.

170. Dastani Z, Hivert M-F, Timpson N, Perry JRB, Yuan X, Scott RA, *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 2012;**8**:e1002607.

171. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;**485**:242–5.

172. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, *et al.* Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet* 2012;**8**:e1002635.

173. Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, *et al.* Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J Med Genet* 2012;**49**:380–5.

174. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, *et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov* 2012;**2**:82–93.

175. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012;**337**:64–9.

176. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012;**380**:572–80.

177. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 2012;**44**:685–9.

178. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* 2012;**485**:502–6.

179. Musunuru K, Romaine SPR, Lettre G, Wilson JG, Volcik KA, Tsai MY, *et al.* Multi-ethnic analysis of lipid-associated loci: the NHLBI CARe project. *PLoS One* 2012;**7**:e36473.

180. Dixon-Salazar TJ, Silhavy JL, Udpa N, Schroth J, Bielas S, Schaffer AE, *et al.* Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med* 2012;**4**:138ra78.

181. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 2012;**486**:405–9.

182. Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A, *et al.* De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet* 2012;**44**:941–5.

183. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, *et al.* A landscape of driver mutations in melanoma. *Cell* 2012;**150**:251–63.

184. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, *et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 2012;**488**:106–10.

185. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 2012;**13**:375.

186. Larman TC, DePalma SR, Hadjipanayis AG, Cancer Genome Atlas Research Network, Protopopov A, Zhang J, *et al.* Spectrum of somatic mitochondrial mutations in five cancers. *Proc Natl Acad Sci U S A* 2012;**109**:14087–91.

187. Friedman J, Olvera J, Silhavy JL, Gabriel SB, Gleeson JG. Mild paroxysmal kinesigenic dyskinesia caused by PRRT2 missense mutation with reduced penetrance. *Neurology* 2012;**79**:946–8.

188. Novarino G, El-Fishawy P, Kayserili H, Meguid NA, Scott EM, Schroth J, *et al.* Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. *Science* 2012;**338**:394–7.

189. Bick AG, Flannick J, Ito K, Cheng S, Vasan RS, Parfenov MG, *et al.* Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *Am J Hum Genet* 2012;**91**:513–9.

190. Manzini MC, Tambunan DE, Hill RS, Yu TW, Maynard TM, Heinzen EL, *et al.* Exome sequencing and functional validation in zebrafish identify GTDC2 mutations as a cause of Walker-Warburg syndrome. *Am J Hum Genet* 2012;**91**:541–7.

191. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;**150**:1107–20.

192. Puffenberger EG, Jinks RN, Wang H, Xin B, Fiorentini C, Sherman EA, *et al.* A homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder. *Hum Mutat* 2012;**33**:1639–46.

193. Haas JT, Winter HS, Lim E, Kirby A, Blumenstiel B, DeFelice M, *et al.* DGAT1 mutation is linked to a congenital diarrheal disorder. *J Clin Invest* 2012;**122**:4680–4.

194. Çalışkan M, Chong JX, Uricchio L, Anderson R, Chen P, Sougnez C, *et al.* Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13. *Hum Mol Genet* 2011;**20**:1285–9.

195. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, *et al.* A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol* 2011;**12**:R1.

196. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, *et al.* The genomic complexity of primary human prostate cancer. *Nature* 2011;**470**:214–20.

197. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, *et al.* Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet* 2011;**7**:e1001300.

198. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011;**471**:467–72.

199. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8.

200. Shea J, Agarwala V, Philippakis AA, Maguire J, Banks E, Depristo M, *et al.* Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. *Nat Genet* 2011;**43**:801–5.

201. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* 2011;**333**:1157–60.

202. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat Genet* 2011;**43**:964–8.

203. Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, *et al.* The functional spectrum of low-frequency coding variation. *Genome Biol* 2011;**12**:R84.

204. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;**43**:1066–73.

205. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 2011;**365**:2497–506.

206. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;**17**:98–110.

207. Jakkula E, Leppä V, Sulonen A-M, Varilo T, Kallio S, Kemppinen A, *et al.* Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am J Hum Genet* 2010;**86**:285–91.

208. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;**463**:899–905.

209. Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, *et al.* Integrative analysis of the melanoma transcriptome. *Genome Res* 2010;**20**:413–27.

210. Musunuru K, Lettre G, Young T, Farlow DN, Pirruccello JP, Ejebe KG, *et al.* Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet* 2010;**3**:267–75.

211. Valente EM, Logan CV, Mougou-Zerelli S, Lee JH, Silhavy JL, Brancati F, *et al.* Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes. *Nat Genet* 2010;**42**:619–25.

212. Blumenstiel B, Cibulskis K, Fisher S, DeFelice M, Barry A, Fennell T, *et al.* Targeted exon sequencing by in-solution hybrid selection. *Curr Protoc Hum Genet* 2010;**Chapter 18**:Unit 18.4.

213. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–303.

214. National Heart, Lung, and Blood Institute working group, Fabsitz RR, McGuire A, Sharp RR, Puggal M, Beskow LM, *et al.* Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. *Circ Cardiovasc Genet* 2010;**3**:574–80.

215. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–13.

216. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;**467**:52–8.

217. Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burtt NP, *et al.* High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet* 2010;**42**:851–8.

218. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010;**11**:685–96.

219. Peloso GM, Demissie S, Collins D, Mirel DB, Gabriel SB, Cupples LA, *et al.* Common genetic variation in multiple metabolic pathways influences susceptibility to low HDL-cholesterol and coronary heart disease. *J Lipid Res* 2010;**51**:3524–32.

220. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, *et al.* Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 2010;**363**:2220–7.

221. Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009;**41**:35–46.

222. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet* 2009;**Chapter 2**:Unit 2.12.

223. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;**27**:182–9.

224. Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009;**41**:334–41.

225. Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, *et al.* De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet* 2009;**41**:931–5.

226. Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* 2009;**69**:7385–92.

227. Patterson N, Gabriel S. Combinatorics and next-generation sequencing. *Nat Biotechnol* 2009;**27**:826–7.

228. Toronto International Data Release Workshop Authors, Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, *et al.* Prepublication data sharing. *Nature* 2009;**461**:168–70.

229. Dutt A, Salvesen HB, Chen T-H, Ramos AH, Onofrio RC, Hatton C, *et al.* Drug-sensitive FGFR2 mutations in endometrial carcinoma. *Proc Natl Acad Sci U S A* 2008;**105**:8713–7.

230. Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008;**40**:1056–8.

231. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 2008;**40**:1166–74.

232. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008;**40**:1253–60.

233. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 2008;**359**:1995–2004.

234. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;**455**:1069–75.

235. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, *et al.* Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One* 2008;**3**:e3583.

236. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008;**40**:204–10.

237. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, *et al.* High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 2007;**39**:347–51.

238. Gabriel S. Population genetic tools: application to cancer. *Semin Oncol* 2007;**34**:S21–4.

239. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;**316**:1331–6.

240. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;**357**:851–62.

241. GAIN Collaborative Research Group, Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Collaborative Association Study of Psoriasis, *et al.* New models of

collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet* 2007;**39**:1045–51.

242. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;**449**:851–61.

243. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;**449**:913–8.

244. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;**450**:893–8.

245. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PIW, Maller J, *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;**39**:1477–82.

246. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, *et al.* Common deletion polymorphisms in the human genome. *Nat Genet* 2006;**38**:86–92.

247. Kathiresan S, Yang Q, Larson MG, Camargo AL, Tofler GH, Hirschhorn JN, *et al.* Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk. *Arterioscler Thromb Vasc Biol* 2006;**26**:1405–12.

248. Meigs JB, Dupuis J, Liu C, O'Donnell CJ, Fox CS, Kathiresan S, *et al.* PAI-1 Gene 4G/5G polymorphism and risk of type 2 diabetes in a population-based sample. *Obesity* 2006;**14**:753–8.

249. Lee JC, Vivanco I, Beroukhim R, Huang JHY, Feng WL, DeBiasi RM, *et al.* Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. *PLoS Med* 2006;**3**:e485.

250. Gabriel S. Variation in the human genome and the inherited basis of common disease. *Semin Oncol* 2006;**33**:S46–9.

251. Litonjua AA, Tantisira KG, Lake S, Lazarus R, Richter BG, Gabriel S, *et al.* Polymorphisms in signal transducer and activator of transcription 3 and lung function in asthma. *Respir Res* 2005;**6**:52.

252. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, *et al.* Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 2005;**308**:107–11.

253. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJP, *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005;**7**:387–97.

254. Kathiresan S, Gabriel SB, Yang Q, Lochner AL, Larson MG, Levy D, *et al.* Comprehensive survey of common genetic variation at the plasminogen activator inhibitor-1 locus and relations to circulating plasminogen activator inhibitor-1 levels. *Circulation* 2005;**112**:1728–35.

255. de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005;**37**:1217–23.

256. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res* 2005;**15**:1576–83.

257. Gabriel S, Ziaugra L. SNP genotyping using Sequenom MassARRAY 7K platform. *Curr Protoc Hum Genet* 2004;**Chapter 2**:Unit 2.12.

258. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, *et al.* Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;**36**:388–93.

259. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;**304**:1497–500.

260. Tantisira KG, Lake S, Silverman ES, Palmer LJ, Lazarus R, Silverman EK, *et al.* Corticosteroid pharmacogenetics: association of sequence variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids. *Hum Mol Genet* 2004;**13**:1353–9.

261. Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, *et al.* Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004;**32**:e71.

262. Sawcer SJ, Maranian M, Singlehurst S, Yeo T, Compston A, Daly MJ, *et al.* Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Hum Mol Genet* 2004;**13**:1943–9.

263. Reich DE, Gabriel SB, Altshuler D. Quality and completeness of SNP databases. *Nat Genet* 2003;**33**:457–8.

264. Baron RM, Palmer LJ, Tantisira K, Gabriel S, Sonna LA, Le L, *et al.* DNA sequence variants in epithelium-specific ETS-2 and ETS-3 are not associated with asthma. *Am J Respir Crit Care Med* 2002;**166**:927–32.

265. Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* 2002;**31**:89–93.

266. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002;**296**:2225–9.

267. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002;**419**:832–7.

268. Weese-Mayer DE, Bolk S, Silvestri JM, Chakravarti A. Idiopathic congenital central hypoventilation syndrome: evaluation of brain-derived neurotrophic factor genomic DNA sequence variation. Am J Med Genet. 2002 Feb 1;107(4):306-10. doi: 10.1002/ajmg.10133. PMID: 11840487.

269. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. Nature. 2001 May 10;411(6834):199-204. doi: 10.1038/35075590. PMID: 11346797.

270. Silverman ES, Liggett SB, Gelfand EW, Rosenwasser LJ, Baron RM, Bolk S, Weiss ST, Drazen JM. The pharmacogenetics of asthma: a candidate gene approach. Pharmacogenomics J. 2001;1(1):27-37. doi: 10.1038/sj.tpj.6500005. PMID: 11913724.

271. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 2000 Sep;26(1):76-80. doi: 10.1038/79216. PMID: 10973253.

272. Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES. SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. Proc Natl Acad Sci U S A. 2000 Oct 24;97(22):12164-9. doi: 10.1073/pnas.210394597. PMID: 11035790; PMCID: PMC17312.

273. Angrist M, Jing S, Bolk S, Bentley K, Nallasamy S, Halushka M, Fox GM, Chakravarti A. Human GFRA1: cloning, mapping, genomic structure, and evaluation as a candidate gene for Hirschsprung disease susceptibility. Genomics. 1998 Mar 15;48(3):354-62. doi: 10.1006/geno.1997.5191. PMID: 9545641.

274. Bolk S, Angrist M, Xie J, Yanagisawa M, Silvestri JM, Weese-Mayer DE, Chakravarti A. Endothelin-3 frameshift mutation in congenital central hypoventilation syndrome. Nat Genet. 1996 Aug;13(4):395-6. doi: 10.1038/ng0896-395. PMID: 8696331.

275. Bolk S, Angrist M, Schwartz S, Silvestri JM, Weese-Mayer DE, Chakravarti A. Congenital central hypoventilation syndrome: mutation analysis of the receptor tyrosine kinase RET. Am J Med Genet. 1996 Jun 28;63(4):603-9. doi: 10.1002/(SICI)1096-8628(19960628)63:4<603::AID-AJMG14>3.0.CO;2-M. PMID: 8826440.

276. Angrist M, Bolk S, Thiel B, Puffenberger EG, Hofstra RM, Buys CH, Cass DT, Chakravarti A. Mutation analysis of the RET receptor tyrosine kinase in Hirschsprung disease. Hum Mol Genet. 1995 May;4(5):821-30. doi: 10.1093/hmg/4.5.821. PMID: 7633441.

277. Puffenberger EG, Kauffman ER, Bolk S, Matise TC, Washington SS, Angrist M, Weissenbach J, Garver KL, Mascari M, Ladda R, et al. Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet. 1994 Aug;3(8):1217-25. doi: 10.1093/hmg/3.8.1217. PMID: 7987295.