UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TEMPUS AI, INC.,

Petitioner,

v.

GUARDANT HEALTH, INC.,

Patent Owner.

Case IPR2025-01434

U.S. Patent 11,149,306

DECLARATION OF MICHAEL L. METZKER, PH.D. IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-29 OF U.S. PATENT 11,149,306

TABLE OF CONTENTS

Page

I. OVERVIEW AND SUMMARY OF OPINIONS 1

II. MY BACKGROUND AND QUALIFICATIONS 2

III. MATERIALS AND INFORMATION CONSIDERED 9

IV. UNDERSTANDING OF THE LAW 18

A. Prior Art 18

B. Level of Ordinary Skill in the Art 19

C. Claim Construction 19

D. Obviousness 20

V. QUALIFICATIONS AND KNOWLEDGE OF ONE OF ORDINARY SKILL IN THE ART 22

VI. CLAIM CONSTRUCTION 22

VII. TECHNICAL BACKGROUND 23

A. DNA Basics 23

B. Human genetics 25

C. Cell-free DNA 26

D. Genetic Variants & Disease 26

1. *Types of genetic variants* 27

2. *Cell type: how genomic variants are acquired* 30

3. *How genetic variants are identified* 31

E. Sequencing technologies 32

1. *General principles* 32

2. *Next generation sequencing (NGS)* 34

3. *Only a subset of NGS sequence reads is generated* 36

F. Amplification and target enrichment of multiple loci 37

1. *Polymerase chain reaction (PCR) amplification* 37

2. *Multiplex PCR* 41

3. *Hybrid capture* 42

G. Library preparation, Including Tagging With Barcodes 45

H. Sequence reads 46

I. Mapping/Alignment 47

J. Generating consensus sequences 51

K. Copy number variant analysis 52

VIII. THE ’306 PATENT 54

A. Overview And Independent Claims 54

B. ’306 Patent Prosecution History 60

IX. THE PRIOR ART References 64

A. PCT Publication No. WO 2013/123442 A1 (“PCT ’442 Appl. (Bielas)”) 64

B. PCT Publication No. WO 2012/142213 A2 (“PCT ’213 Appl. (Vogelstein)”) 68

C. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA (“Forshew (2012)”) 69

D. PCT Publication No. WO 2012/099832 A2 (“PCT ’832 Appl. (Hendricks)”) 70

E. U.S. Patent Publication No. 2014/0296081 (the “’081 Patent Appl. (Diehn)”) 71

F. U.S. Patent No. 9,404,156 (the “’156 Patent (Hicks)”) 72

X. OPINIONS REGARDING THE VALIDITY OF THE ’306 PATENT CLAIMS 73

A. Ground 1: Claims 1-2, 7-8, 12-13, 17-18, 20-21, 24-26, and 29 Are Rendered Obvious By PCT ’442 Appl. (Bielas) in View of PCT ’213 Appl. (Vogelstein) 73

1. Motivation To Combine 74

2. Independent Claim 1 78

3. Claim 2: “The method of claim 1, wherein the population of cfDNA molecules is obtained or derived from a subject having cancer.” 93

4. Claim 7: “The method of claim 1, wherein z is between 2 and 8.” 95

5. Claim 8: “The method of claim 1, wherein the molecular barcodes have a length of 5 to 20 base pairs.” 96

6. Claim 12: “The method of claim 1, wherein the duplex tags are part of sequencing adapters.” 97

7. Claim 13: “The method of claim 1, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises grouping the paired reads or the unpaired reads into families based at least in part on (i) the molecular barcodes associated withthe paired reads or the unpaired reads and (ii) sequence information at the start and stop positions of the paired reads or the unpaired reads.” 99

8. Claim 17 101

9. Claim 18: “The method of claim 17, wherein the sample is blood, plasma, or serum.” 109

10. Claim 20: “The method of claim 17, wherein the tagging comprises ligating the molecular barcodes to double-stranded cfDNA molecules.” 109

11. Claim 21: "The method of claim 17, wherein the set of tags comprises 2 to 10,000 different molecular barcode sequences." 110

12. Claim 24: "The method of claim 17, wherein (d) further comprises reducing or tracking redundancy of a plurality of sequence reads from the set of sequencing reads, wherein the reducing or tracking comprises mapping at least a subset of the plurality of equence reads to a reference sequence, and the reducing or tracking is based on (i) the molecular barcodes associated the paired reads or the unpaired reads and (ii) sequence information at the start and stop positions of the paired reads or the unpaired reads." 111

13. Claim 25: "The method of claim 24, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises determining a base call at one or more genomic loci for a plurality of sequence reads that map to the one or more genomic loci onthe reference sequence." 112

14. Claim 26: "The method of claim 24, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises collapsing the plurality of sequence reads to produce consensus sequences representative of a sequence of the original double-strnded cfDNA molecules from among the tagged parent polynucleotides." 114

15. Claim 29: "The method of claim 1, wherein the distinct cfDNA molecules in (e) are determined based on (i) the paired reads and (ii) the unpaired reads." 115

B. Ground 2: Claims 3, 9, 10, 11, 19 and 22-23 Are Rendered Obvious By PCT '442 Appl. (Bielas) in View of PCT '213 Appl. (Vogelstein) and Forshew (2012) 117

1. Motivation To Combine 117

2. **Claim 3: "The method of claim 1, wherein the population of cfDNA molecules comprises 1 nanogram (ng) to 100 ng of cfDNA molecules."** 120

3. **Claim 9: "The method of claim 1, further comprising, prior to the sequencing, enriching at least a subset of the amplified progeny polynucleotides for target regions of interest to produce enriched progeny polynucleotides."** 121

4. **Claim 10: "The method of claim 9, wherein the target regions of interest comprise genetic sequences of a plurality of genes selected from the group consisting of ALK, APC, BRAF, CDKN2A, EGFR, ERBB2, FBXW7, KRAS, MYC, NOTCH1, NRAS, PIK3CA, PTEN, RB1, TP5, MET, AR, ABL1, AKT1, ATM, CDH1, CSF1R, CTNNB1, ERBB4, EZH2, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAQ, GNAS, HNF1A, HRAS, IDH1, IDH2, JAK2, JAK3, KDR, KIT, MLH1, MPL, NPM1, PDGFRA, PROC, PTPN11, RET, SMAD4, SMARCB1, SMO, SRC, STK11, VHL, TERT, CCND1, CDK4, CDKN2B, RAF 1, BRCA1, CCND2, CDK6, NF1, TP53, ARID1A, BRCA2, CCNE1, ESR1, RIT1, GATA3, MAP2K1, RHEB, ROS1, ARAF, MAP2K2, NFE2L2, RHOA, and NTRK1."** 123

5. **Claim 11: "The method of claim 10, further comprising, prior to the sequencing, amplifying a plurality of the enriched progeny polynucleotides."** 124

6. **Claim 19: "The method of claim 17, wherein the population of double-stranded cfDNA molecules comprises 1 nanogram (ng) to 100 ng of double-stranded cfDNA molecules."** 125

7. **Claim 22: "The method of claim 17, further comprising, prior to the sequencing, enriching at least a subset of the amplified progeny polynucleotides for target regions of interest to produce enriched progeny polynucleotides."** 126

8. **Claim 23: "The method of claim 22, wherein the target regions of interest comprise genetic sequences of a plurality of genes selected from the group consisting of ALK, APC, BRAF, CDKN2A, EGFR, ERBB2, FBXW7, KRAS, MYC, NOTCH1, NRAS, PIK3CA, PTEN, RB1, TP53, MET, AR, ABL1, AKT1, ATM, CDH1, CSF1R, CTNNB1, ERBB4, EZH2, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAQ, GNAS, HNF1A, HRAS, IDH1, IDH2, JAK2, JAK3, KDR, KIT, MLH1, MPL, NPM1, PDGFRA, PROC, PTPN11, RET, SMAD4, SMARCB1, SMO, SRC, STK11, VHL, TERT, CCND1, CDK4, CDKN2B, RAF1, BRCA1, CCND2, CDK6, NF1, TP53, ARID1A, BRCA2, CCNE1, ESR1, RIT1, GATA3, MAP2K1, RHEB, ROS1, ARAF, MAP2K2, NFE2L2, RHOA, and NTRK1."** 126

C. Ground 3: Claims 4, 5 and 6 Are Rendered Obvious By PCT '442 Appl. (Bielas) in View of PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks) and/or '081 Patent Appl. (Diehn) 127

1. **Motivation to Combine** 127

2. **Claim 4: "The method of claim 1, wherein the molecular barcodes are ligated to the plurality of the cfDNA molecules using more than a 10× excess of duplex tags as compared to the population of cfDNA molecules, wherein at least 20% of the cfDNA molecules from the population are tagged with the duplex tags."** 130

3. **Claim 5: "The method of claim 1, wherein the molecular barcodes are ligated to the plurality of the cfDNA molecules using more than an 80× excess of duplex tags as compared to the population of cfDNA molecules."** 132

4. **Claim 6: "The method of claim 5, wherein at least 40% of the cfDNA molecules from the population are tagged with the duplex tags."** 133

D. Ground 4: Claims 14-16 and 27-28 Are Rendered Obvious By PCT ’442 Appl. (Bielas) in View of PCT ’213 Appl. (Vogelstein) and the ’156 Patent (Hicks) 133

1. **Motivation to Combine** 133

2. **Claim 14: “The method of claim 1, further comprising: (f) determining quantitative measures of (i) paired reads or (ii) unpaired reads that map to a genomic locus of the reference sequence.”** 135

3. **Claim 15: “The method of claim 14, further comprising: (g) estimating with a programmed computer processor a quantitative measure of tagged parent polynucleotides that map to the genomic locus based on the quantitative measures of the paired reads and he unpaired reads.”** 137

4. **Claim 16: “The method of claim 15, further comprising: (h) detecting copy number variation in the population of cfDNA molecules by determining a normalized quantitative measure determined in (g) at each of one or more genomic loci.”** 139

5. **Claim 27: “The method of claim 24, further comprising determining quantitative measures of at least two of (i) paired reads that map to one or more genomic loci, (ii) unpaired reads that map to one or more genomic loci, (iii) a read depth of the paired eads, and (iv) a read depth of the unpaired reads.”** 139

6. **Claim 28: “The method of claim 27, further comprising estimating with a programmed computer processor a quantitative measure of tagged parent polynucleotides that map to a genomic locus of the reference sequence based on the quantitative measures of at east two of (i) paired reads, (ii) unpaired reads, (iii) a read depth of the paired reads, and (iv) a read depth of the unpaired reads at the genomic locus.”** 141

XI. Secondary Considerations Cannot Rebut the Strong Showing of Obviousness 144

I, Michael L. Metzker, hereby declare as follows.

1. I am over the age of eighteen (18) and otherwise competent to make this declaration.

2. I have been retained as an expert witness on behalf of Tempus AI, Inc. ("Tempus") in connection with the above-captioned requested *inter partes* review ("IPR").

I. OVERVIEW AND SUMMARY OF OPINIONS

3. I understand that the petition for *inter partes* review involves Claims 1-29 of the '306 Patent[1], which issued on October 19, 2021.[2] I have been informed by counsel and understand that the '306 Patent is assigned to Guardant Health, Inc. ("Guardant").

4. I have been informed by counsel and understand that the '306 Patent resulted from U.S. Patent Application No. 16/945,124 ("the '124 Application"), filed on July 31, 2020, naming AmirAli Talasaz, Helmy Eltoukhy, and Stefanie Ann Ward Mortimer as inventors.[3]

5. I have been informed by counsel and understand that the '306 Patent

---

[1] *See* Ex. 1001, Talasaz *et al.*, *Methods and systems for detecting genetic variants*, U.S. Patent 11,149,306 (2021) (the "'306 Patent").

[2] *Id*. at cover page.

[3] *Id*.

also claims priority to U.S. Provisional Patent Application No. 61/948,509 ("the '509 Provisional"), filed on Mar. 5, 2014, and U.S. Provisional Patent Application No. 61/921,456 ("the '456 Provisional"), filed on Dec. 28, 2013.[4]

6. It is my opinion that Claims 1-29 of the '306 Patent are unpatentable as obvious in view of the prior art. My opinion is based on the following grounds:

Ground	Description
Ground 1	Claims 1-2, 7-8, 12-13, 17-18, 20-21, 24-26, and 29 are rendered obvious under 35 U.S.C. § 103 by PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein)
Ground 2	Claims 3, 9, 10, 11, 19, and 22-23 are rendered obvious under 35 U.S.C. § 103 by PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein) and Forshew (2012)
Ground 3	Claims 4, 5 and 6 are rendered obvious under 35 U.S.C. § 103 by PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks) and the '081 Patent Appl. (Diehn)
Ground 4	Claims 14-16 and 27-28 are rendered obvious under 35 U.S.C. § 103 by PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks)

II. MY BACKGROUND AND QUALIFICATIONS

7. I am currently Founder, President, and Chief Executive Officer of RedVault Biosciences, LP ("RedVault"), which I founded in April 2013. RedVault is a biotechnology company committed to creating innovative technologies to advance genomic medicine. Before taking my position at RedVault, I was a tenured

[4] *Id.*

and adjunct Associate Professor in both the Department of Molecular & Human Genetics and at the Human Genome Sequencing Center ("HGSC") at Baylor College of Medicine ("BCM"). I have also been an adjunct Associate Professor in the Cell & Molecular Biology Program at BCM and adjunct Assistant and Associate Professor in Chemistry at Rice University. At BCM, I worked as part of the team at the HGSC that contributed to the Human Genome Project.

8. From September 2022 to September 2023, I was also Co-Founder and Chief Technical Officer of 454 Bio, Inc. 454 Bio is developing a portable DNA sequencer, a customizable device for decentralized, on-site sequencing.

9. In March 2002, I founded LaserGen, Inc. ("LaserGen"), which provides technology and reagents used in, among other things, nucleic acid sequencing methods. From March 2002 to April 2012, I was President and Chief Executive Officer of LaserGen, and from April 2012 to October 2012, I held the position of Chief Technology Officer. I was also a member of the Board of Directors from March 2002 to October 2012. From 2016 to 2018, Agilent Technologies acquired LaserGen for $185M to convert its reversible terminator technology into a clinical diagnostic platform. I currently have no working relationship with LaserGen.

10. Since 1988, I have performed scientific research in the fields of molecular biology and chemical synthesis of nucleoside and nucleotide inhibitors. In particular, I have extensive experience with technology development in DNA

sequencing methods including genomic DNA isolation, polymerase chain reaction ("PCR"), fragmenting genomic DNA, genomic DNA and complementary DNA ("cDNA") library construction, bacterial cloning, nucleic acids chemistry, DNA modifying enzymes, polymerase engineering, fluorescent dyes, fluorescence imaging, and data analysis of multi-color images. I also have extensive experience in the use and development of methods utilized to detect sequence variation, including single nucleotide variants ("SNVs") and single nucleotide polymorphisms ("SNPs") in all organisms, with a particular focus in humans and human immunodeficiency virus ("HIV").

11. My research has been devoted in large part to developing next-generation sequencing ("NGS") technologies, developing novel methods to study HIV transmission between individuals, and identifying molecular causes of novel forms of diabetes and their treatment. My colleagues and I have been deeply involved in PCR, DNA fragmentation, library construction, cloning, and the development of sequencing technologies. My industrial experience entails performing analytical chemical testing of inorganic materials (Aerojet-General), organic synthesis and characterization of porous polystyrene beads for applications in high performance liquid chromatography (Bio-Rad), development and applications of a robotic platform in Sanger sequencing (Applied Biosystems), and development and applications of high-throughput DNA sequencing to identify novel

gene targets for small molecule screening (Merck Research Laboratories).

12. I received my Bachelor of Science degree in Biochemistry & Biophysics from the University of California, Davis in 1984. Between 1985 and 1988, I enrolled in a Master of Science program in the department of Chemistry at San Francisco State University. Although I completed the required coursework for a Master of Science degree with an emphasis in Organic Chemistry, I elected not to complete this degree. Instead, in 1991, I enrolled in a doctoral program in the Department of Molecular and Human Genetics at Baylor College of Medicine in Houston, Texas. In 1996, I received my Ph.D. in Molecular and Human Genetics.

13. From 2000 to 2018, I gave lectures on sequencing technologies to graduate and medical students as part of the *Molecular Methods* course at BCM. These lectures were titled, *cDNA and Genomic Libraries, First-generation Sequencing, and Genotyping*, and *Next-generation Sequencing*. From 2001 to 2003, I also gave a lecture titled, *Mammalian Genome Analysis* in the *Mammalian Genomics* course at BCM.

14. I have also spoken at a variety of academic and industry conferences in the fields of molecular biology and DNA sequencing, including the *Next-Generation Sequencing* Workshop at Lübeck University in Germany; the *Centre de Regulació Genòmica* ("CRG") Symposium, Barcelona, Spain; the *Next-Generation Sequencing* Conference in Boston, MA, the *Advances in Genome Biology and*

Technology Meeting ("AGBT") in Marco Island, FL; the *Copenhagenomics* Meeting in Copenhagen, Denmark; the *American Society of Microbiology* Conference in San Francisco, CA; the SelectBio's *Advances in Next Generation Sequencing* Meeting; and the *Critical Path to TB Drug Regimens* ("CPTR") Workshop in Washington, DC.

15. I have authored 58 peer-reviewed papers and seven book chapters. Several of my review papers discuss emerging technologies and advances in DNA sequencing, including detection of sequence variations. I have written book chapters on topics related to polymerase chain reaction ("PCR") and DNA sequencing. I have also reviewed numerous manuscripts for peer-review publications in the fields of molecular biology and DNA sequencing, including manuscripts applying the sequencing techniques described *supra*.

16. I have also conducted extensive research in the fields of molecular biology and DNA sequencing. My research has investigated and advanced fundamental techniques such as Sanger sequencing and NGS methods, chemical synthesis of nucleoside and nucleotide inhibitors, Sanger sequencing and NGS approaches for HIV forensics, and detection of genomic variation. Some of this research involved developing automated Sanger sequencing assays to measure mixed nucleotide ratios in DNA sequences of interest. In addition, my colleagues and I developed a novel diagnostic platform using intracellular and extracellular

microRNAs for detection of various analytes. Like cell-free DNA, extracellular microRNAs are small nucleic acids that can be used as biomarkers for diagnostic purposes. My research has been funded by the National Institutes of Health ("NIH"), National Institute of Justice ("NIJ"), United States Department of Agriculture ("USDA"), and private foundations. Since 1999, I have received or have been associated with numerous grants largely focusing on gene and genome sequencing totaling over $170 million.

17. I am an inventor on 57 U.S. and European issued patents and patent applications directed towards molecular biology, DNA sequencing, and chemical synthesis. For example, I coinvented labelled and unlabeled cleavable terminating groups and methods for DNA sequencing described in U.S. Patent No. 8,148,503; 9,200,319 and 10,041,115. I also coinvented methods for the use of BODIPY fluorophore-labeled DNA for dye-primer sequencing described in U.S. Patent No. 5,861,287 and oligonucleotides labelled with BODIPY fluorophore compounds described in U.S. Patent No. 5,994,063. Further, I coinvented the pulsed-multiline excitation ("PME") method in collaboration with the late Nobel Laureate Robert F. Curl, Ph.D., which resulted in U.S. Patent Nos. 6,995,841; 7,511,811; and 8,089,628. By way of collaboration with my scientific team at LaserGen, I also coinvented novel nucleotide terminators, which resulted in numerous patents including U.S. Patent Nos. 7,893,227; 7,897,737; 7,964,352; 8,148,503; 8,198,029; 8,361,727;

8,497,360; 8,877,905; 8,889,860; 8,969,535; 9,200,319; 9,399,798; 9,689,035; 10,041,115; and 11,001,886.

18. I have served on review panels for Genome Canada, Canadian Institute for Health Research ("CIHR"), National Aeronautics and Space Administration ("NASA"), and the U.S. Department of Energy ("DOE"), as well as various NIH study sections for the National Cancer Institute ("NCI"), National Institute of Biomedical Imaging and Bioengineering ("NIBIB"), National Human Genome Research Institute ("NHGRI"), and National Institute for Allergy and Infectious Diseases ("NIAID"). In March 2018, I became a member of the College of Reviewers for CIHR.

19. From 2003 to 2006, I served on the editorial/advisory board for the journal *Genome Research*. From 2006 to 2012, I served as a scientific organizer for the AGBT Meeting. From 2011 to 2013, I served on the advisory committee of Genome Canada: *Advancing Technology Innovation Through Discovery* ("ATID") projects – *Finding of Rare Disease Genes* in Canada ("FORGE Canada"), and the *Canadian Pediatric Cancer Genome Consortium* ("CPCGC").

20. I have also served as an expert witness in several criminal trials in the United States involving the transmission of human immunodeficiency virus ("HIV") between individuals. The criminal case involving Richard Schmidt was portrayed in an episode titled *Shot of Vengeance* on Forensic Files in 2003. The criminal case

involving Philippe Padieu was reported on ABC News 20/20 by Elizabeth Vargas and on Oprah, both in 2009, and helped inspired an episode titled *Quickie* on Law & Order: Special Victims Unit in 2010.

21. I belong to the American Association for the Advancement of Science, the American Chemical Society, and the Texas Genetics Society. My professional experience, education, and publications are presented in greater detail in my curriculum vitae (Ex. 1004), including publications and patents authored in the previous five years, as well as cases in which I provided expert testimony at deposition or trial.

22. Accordingly, I am an expert in molecular biology and sequencing technologies, including nucleic acid library preparation, Sanger sequencing and next-generation sequencing, assay development and detection and analysis of labeled nucleotides, and have been since prior to December 28, 2013. For that reason, I am qualified to provide an opinion as to what one of ordinary skill in the art would have understood, known, or concluded before December 28, 2013.

23. I am being compensated for my time in connection with this IPR at my standard consulting rate, which is $750 per hour. My compensation does not depend in any way on the opinions I express or the outcome of this IPR.

III. MATERIALS AND INFORMATION CONSIDERED

24. In formulating my opinions, I have relied upon my 30+ years of

education, training, knowledge, and experience in the relevant art. In formulating my opinions, I have also considered the viewpoint of one of ordinary skill in the art ("POSA")[5] prior to December 28, 2013.

25. In preparing this Declaration, I have reviewed the '306 Patent, the file history of the '306 Patent, and the materials cited herein, including the following exhibits, in light of the general knowledge in the art before December 28, 2013. The materials cited herein are publications that were typically read and reasonably relied upon by one of ordinary skill in this field, including the bibliographic information presented therein.

Exhibit	Description
1001	Talasaz *et al.*, *Methods and systems for detecting genetic variants*, U.S. Patent No. 11,149,306 (2021) ("the '306 Patent")
1002	Prosecution history for the '306 Patent ("'306 Patent File History")
1003	Declaration of Michael Metzker, Ph.D.
1004	Curriculum Vitae of Michael Metzker, Ph.D.
1005	Bielas *et al.*, *Compositions and methods for accurately identifying mutations*, PCT Publication No. WO 2013/123442 A1 (2013) (the "PCT '442 Appl. (Bielas)")
1006	Vogelstein *et al.*, *Safe sequencing system*, PCT Publication No. WO 2012/142213 A2 (2012) (the "PCT '213 Appl. (Vogelstein)")

[5] Throughout this Report, I also refer to one of ordinary skill in the art as a POSA.

1007	Fodor *et al.*, *Digital counting of individual molecules by stochastic attachment of diverse labels*, U.S. Patent Application Publication No. 2011/0160078 (2011) (the "'078 Patent Appl. (Fodor)")
1008	Hendricks, *Enzymatic ligation of nucleic acids*, PCT International Publication No. WO 2012/099832 A2 (2012) (the "PCT '832 Appl. (Hendricks)")
1009	Diehn *et al.*, *Identification and use of circulating tumor markers*, U.S. Patent Application Publication No. 2014/0296081 (2014) (the "'081 Patent Appl. (Diehn)")
1010	Forshew *et al.*, *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA*, SCIENCE TRANSLATIONAL MEDICINE 4:136ra68, pp. 1–12 (2012) ("Forshew (2012)")
1011	Hicks *et al.*, *Varietal counting of nucleic acids for obtaining genome copy number information*, U.S. Patent No. 9,404,156 (the "'156 Patent (Hicks)")
1012	K. Shiroguchi *et al.*, *Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes*, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES 109:1347–1352 (2012) ("Shiroguchi (2012)")
1013	Saxonov, *Methods and compositions for nucleic acid analysis*, U.S. Patent Application Publication No. 2012/0316074 A1 (the "'074 Patent Appl. (Saxonov)")
1014	Schwarzenbach *et al.*, *Cell-free nucleic acids as biomarkers in cancer patients*, NATURE REVIEWS CANCER 11:426–437 (2011) ("Schwarzenbach (2011)").
1015	Human genome variation, fact sheet, National Human Genome Research Institute (NHGRI) at https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-

	variation#:~:text=It%20is%20frequently%20stated%20that%20any%20two%20peoples%E2%80%99,peoples%E2%80%99%20genomes%20once%20every%201%2C300%20nucleotides%20or%20so (last visited July 25, 2025) ("NHGRI variation fact sheet")
1016	*Guardant Health, Inc. v. Foundation* Medicine, IPR2017-01448, Paper 2 – Petition for Inter Partes Review of U.S. Patent No. 9,340,830
1017	Metzker, *Sequencing technologies — the next generation*, NATURE REVIEWS GENETICS 11:31–46 (2010) ("Metzker (2010)")
1018	*Guardant Health, Inc. v. University of Washington*, IPR2022-00816, Paper 3 – Petition for Inter Partes Review of U.S. Patent No. 10,760,127
1019	Metzker & Caskey, *Polymerase chain reaction*, In ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION, Second Edition, Volume 5 (2006) ("Metzker (2006)").
1020	Mamanova *et al.*, *Target-enrichment strategies for next-generation sequencing*, NATURE METHODS 7:111–118 (2010) ("Mamanova (2010)")
1021	*Guardant Health, Inc. v. University of Washington*, IPR2022-00450, Paper 3 – Petition for Inter Partes Review of U.S. Patent No. 10,689,699
1022	*Guardant Health, Inc. v. University of Washington*, IPR2022-01388, Paper 2 – Petition for Inter Partes Review of U.S. Patent No. 10,689,699
1023	[Intentionally Omitted]
1024	*Guardant Health, Inc. v. University of Washington*, IPR2022-00935, Paper 2 – Petition for Inter Partes Review of U.S. Patent No. 10,287,631
1025	U.S. Provisional Patent Application No. 61/600535 ("Diehn Provisional")

1026	*Twinstrand Biosciences, Inc. v. Guardant Health, Inc.*, IPR2022-01400, Paper 2 – Petition for Inter Partes Review of U.S. Patent No. 11,149,306
1027	*Twinstrand Biosciences, Inc. v. Guardant Health, Inc.*, IPR2022-01400, Paper 41 – Final Written Decision
1028	*Twinstrand Biosciences, Inc. v. Guardant Health, Inc.*, IPR2022-01400, Paper 9 – Institution Decision
1029	Li & Weeks, *Structure-independent and quantitative ligation of single-stranded DNA*, ANALYTICAL BIOCHEMISTRY (2005) ("Li (2005)")
1030	Chee, *Nucleic acid constructs and methods of use*, U.S. Patent No. 9,085,798 (2015) (the "'798 Patent (Chee)")
1031	Thomas *et al.*, *Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing*, NATURE MEDICINE 12:852–855 (2006) ("Thomas (2006)")
1032	Buckingham, *Chromosomal structure and chromosomal mutations*, In MOLECULAR DIAGNOSTIC FUNDAMENTAL, METHODS, & CLINICAL APPLICATIONS, Eds. Buckingham & Flaws, F.A. Davis Company, Chapter 8, pp. 155–172 (2007) ("Buckingham (2007)")
1033	Gemayel *et al.*, *Variable tandem repeats accelerate evolution of coding and regulatory sequences*, ANNUALS REVIEW OF GENETICS 44:445–477 (2010) ("Gemayel (2010)")
1034	Tóth *et al.*, *Microsatellites in different eukaryotic genomes: Survey and analysis*, GENOME RESEARCH 10:967–981 (2000) ("Tóth (2000)")
1035	Laghi *et al.*, *Differences and evolution of the methods for the assessment of microsatellite instability*, ONCOGENE 27:6313–6321 (2008) ("Laghi (2008)")

1036	Richard & Pâques, *Mini- and microsatellite expansions: the recombination connection*, EMBO REPORTS 1:122–126 (2000) (“Richard (2000)”)
1037	Hastings *et al.*, *Mechanisms of change in gene copy number*, NATURE REVIEWS GENETICS 10:551–564 (2009) (“Hastings (2009)”)
1038	Hiatt *et al.*, *Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation*, GENOME RESEARCH 23:843–854 (2013) (“Hiatt (2013)”)
1039	Somatic & germline mutations, https://my.clevelandclinic.org/health/body/23067-somatic--germline-mutations (last visited July 30, 2025) (“Cleveland Clinic webpage”)
1040	Gene changes and cancer, https://www.cancer.org/cancer/understanding-cancer/genes-and-cancer/gene-changes.html (last visited July 30, 2025) (“American Cancer Society webpage”)
1041	International Human Genome Sequencing Consortium, *Initial sequencing and analysis of the human genome*, NATURE 409:860–921 (2001) (“IHGSC (2001)”)
1042	International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome*, NATURE 431:931–945 (2004) (“IHGSC (2004)”)
1043	Forbes, *et al.*, *COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer*, NUCLEIC ACIDS RESEARCH 39:D945–D950 (2011) (“Forbes (2011)”)
1044	*Genomic Data Commons Data Portal*, https://portal.gdc.cancer.gov/ ((last visited July 30, 2025) (“GDC data portal”)
1045	Metzker, *Emerging technologies in DNA sequencing*, GENOME RESEARCH 15:1767–1776 (2005) (“Metzker (2005)”)

1046	Rothberg *et al.*, *An integrated semiconductor device enabling non-optical genome sequencing*, NATURE 475:348-352 (2011) ("Rothberg (2011)")
1047	Jain *et al.*, *Improved data analysis for the MinION nanopore sequencer*, NATURE METHODS 12:351-356 (2015) ("Jain (2015)")
1048	Kinde *et al.*, *Detection and quantification of rare mutations with massively parallel sequencing*, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES 108:9530-9535 (2011) ("Kinde (2011)")
1049	Turner *et al.*, *Methods for genomic partitioning*, ANNUAL REVIEW OF GENOMICS AND HUMAN GENETICS 10:263-284 (2009) ("Turner (2009)")
1050	Edwards & Gibbs, *Multiplex PCR: Advantages, development, and applications*, GENOME RESEARCH 3:S65-S75 (1994) ("Edwards (1994)")
1051	Sharma *et al.*, *(TG/CA)n repeats in human gene families: abundance and selective patterns of distribution according to function and gene length*, BMC GENOMICS 6:83 pp. 1–12 (2005) ("Sharma (2005)")
1052	Gnirke *et al.*, *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*, NATURE BIOTECHNOLOGY 27:182–189 (2009) ("Gnirke (2009)").
1053	Meyerson *et al.*, *Advances in understanding cancer genomes through second-generation sequencing*, NATURE REVIEWS GENETICS 11:685–696 (2010) ("Meyerson (2010)")
1054	Preston *et al.*, *Innovation at Illumina: The road to the $600 human genome*, NATURE PORTFOLIO (2023) at https://www.nature.com/articles/d42473-021-00030-9 ("Preston (2023)").

1055	Bentley *et al.*, *Accurate whole human genome sequencing using reversible terminator chemistry*, NATURE 456:53–59 (2008) ("Bentley (2008)")
1056	Kircher *et al.*, *Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform*, NUCLEIC ACIDS RESEARCH 40:e3 pp. 1–8 (2012) ("Kircher 2012)")
1057	Pierce *et al.*, *A unique and universal molecular barcode array*, NATURE METHODS 3:601–603 (2006) ("Pierce (2006)")
1058	Schmitt *et al.*, *Detection of ultra-rare mutations by next-generation sequencing*, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES 109: 14508-14513 (2012) ("Schmitt (2012)")
1059	Glenn, *Field guide to next-generation DNA sequencers*, MOLECULAR ECOLOGY RESOURCES 11:759-769 (2011) ("Glenn (2011)")
1060	Cock *et al.*, *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants*, NUCLEIC ACIDS RESEARCH 38:1767–1771 (2010) ("Cock (2010)")
1061	CASAVA 1.8: enhanced variant calling in whole-genome resequencing data (2011) ("CASAVA User Guide")
1062	Li *et al.*, *Mapping short DNA sequencing reads and calling variants using mapping quality scores*, GENOME RESEARCH 18:1851–1858 (2008) ("Li (2008)")
1063	Li & Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform*, BIOINFORMATICS 25:1754–1760 (2009) ("Li & Durbin (2009)")
1064	Li & Durbin, *Fast and accurate long-read alignment with Burrows–Wheeler transform*, BIOINFORMATICS 26: 589–595 (2010) ("Li & Durbin (2010)")
1065	Langmead *et al.*, *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*, GENOME BIOLOGY 10:R25.1–R25.10 (2009) ("Langmead (2009)")

1066	Li *et al.*, *The sequence alignment/map format and SAMtools*, BIOINFORMATICS 25:2078–2079 (2009) ("Li (2009)")
1067	DePristo *et al.*, *A framework for variation discovery and genotyping using next-generation DNA sequencing data*, NATURE GENETICS 43:491–498 (2011) ("DePristo (2011)")
1068	McKenna *et al.*, *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data*, GENOME RESEARCH 20:1297-1303 (2010) ("McKenna (2010)")
1069	https://github.com/broadinstitute/gatk/releases (last visited Aug 11, 2025) ("GATK Updates")
1070	Koboldt *et al.*, *VarScan: variant detection in massively parallel sequencing of individual and pooled samples*, BIOINFORMATICS 25:2283–2285 (2009) ("Koboldt (2009)")
1071	Koboldt *et al.*, *VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing*, GENOME RESEARCH 22:568–576 (2012) ("Koboldt (2012)")
1072	Cibulskis *et al.*, *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples*, NATURE BIOTECHNOLOGY 31:213–219 (2013) ("Cibulskis (2013)")
1073	Schmitt *et al. Methods of lowering the error rate of massively parallel DNA sequencing using duplex consensus sequencing*, PCT International Publication No. WO 2013/142389 (2013) (the "'389 PCT Appl. (Schmitt)")
1074-1085	[Intentionally Omitted]
1086	Xi *et al.*, *Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion*, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES 108: E1128–E1136 (2011) ("Xi (2011)")

1087	Quail *et al.*, *Improved protocols for the Illumina Genome Analyzer sequencing system*, CURRENT PROTOCOLS IN HUMAN GENETICS 18.2.1-18.2.27 (2009) (“Quail (2009)”)
1088	Fernández-Suárez & Galperin, The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection, Nucleic Acids Research 41:D1-D7 (2013) (“Fernández-Suárez (2013)”)
1089	Danecek *et al.*, *The variant call format and VCFtools*, BIOINFORMATICS 27:2156–2158 (2011) (“Danecek (2011)”)
1090	Schatz *et al.*, *Assembly of large genomes using second-generation sequencing*, GENOME RESEARCH 20:1165–1173 (2010) (“Schatz (2010)”)
1091	Cooper *et al.*, *Mutational and selective effects on copy-number variants in the human genome*, NATURE GENETICS 39:S22–S29 (2007) (“Cooper (2007)”)

IV. UNDERSTANDING OF THE LAW

26. I am not a lawyer, but I have been informed by counsel and understand the following concerning the applicable law.

A. Prior Art

27. I have been informed by counsel and understand that the law provides certain categories of information, known as prior art, that may be used to render patent claims anticipated or obvious.

28. I have been asked by counsel to assume that the relevant time for my analysis of the pertinent art is on or before December 28, 2013, the earliest claimed priority date for the ’306 Patent.

B. Level of Ordinary Skill in the Art

29. I understand that the assessment of the patentability of the claims of the '306 Patent must be undertaken from the perspective of one of ordinary skill in the art as of the earliest priority date of the '306 Patent, which I have been informed by counsel is December 28, 2013.

30. I have been informed by counsel and understand that the following factors are relevant to determining the level of ordinary skill in the art: (*i*) type of problems encountered in the art, (*ii*) prior art solutions to those problems, (*iii*) rapidity with which innovations are made, (*iv*) sophistication of the technology, and (v) educational level of active workers in the field.

31. I have been informed by counsel and understand that one of ordinary skill in the art is assumed to be aware of all pertinent prior art, and that the skilled artisan has the ability to understand the technology, draw inferences, and make modest adaptations or advances. One of ordinary skill in the art is thus a person of ordinary creativity, not an automaton, and uses common sense.

C. Claim Construction

32. I have been informed by counsel and understand that the first step in comparing prior art to patent claims is to properly construe the claims to determine claim scope and meaning. I also understand that in IPR proceedings, claim terms are presumed to take on the ordinary and customary meaning that they would have

to one of ordinary skill in the art in the context of the entire disclosure of the specification and prosecution history. I further understand that if an inventor acts as his or her own lexicographer, the definition must be set forth in the specification with reasonable clarity, deliberateness, and precision.

D. Obviousness

33. I have been informed by counsel and understand that a patent claim will be rendered obvious and therefore unpatentable if the differences between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to one of ordinary skill in the pertinent art. For the purposes of obviousness, I understand that one of ordinary skill in the art may rely on a single prior art reference or multiple references in combination.

34. I have been informed by counsel and understand that the following factors are considered when determining whether a patent claim would have been obvious to one of ordinary skill in the art: (*i*) the scope and content of the prior art, (*ii*) the differences between the prior art and the claims, (*iii*) the level of ordinary skill in the pertinent art, and (*iv*) the existence of secondary considerations of nonobviousness. These secondary considerations, which I understand are also called "objective indicia" may include factors such as: (*i*) the invention's satisfaction of a long-felt unmet need in the art; (*ii*) unexpected results of the invention; (*iii*)

skepticism of the invention by experts; (*iv*) commercial success of an embodiment of the invention; and (*v*) praise by others for the invention. I also understand that there must be a nexus or connection between the evidence that is the basis for an asserted secondary consideration and the scope of the claimed invention.

35. I have been informed by counsel and understand that a claimed invention can be rendered obvious when, for example, there is some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill in the art to modify the prior art reference or to combine prior art reference disclosure to arrive at the claimed invention. In other words, even if one reference does not show the whole of the invention, if it would have been obvious to one of ordinary skill in the art at the relevant time to add the missing pieces to the invention (for example as a matter of standard practice or application of a well-known principle in the field), then a single reference can render a claim invalid even if it does not show the whole invention. Moreover, a combination of two or more prior art references can render a claim invalid as obvious whether or not there is an explicit suggestion in one of the references to combine the two references, if as a matter of skill or practice in the field, it would be routine to do so. Further, I understand that obviousness does not require absolute predictability. Only a reasonable expectation that the beneficial result will be achieved by the modifications or combination is necessary to show obviousness. I also understand that a patent may be invalid as

obvious to try if the options that one of ordinary skill would have encountered were finite, small, or easily traversed, and one of ordinary skill in the art would have had a reason to select the route that produced the claimed invention. I understand that this motivation from the art can be explicit or implicit using a skilled artisan's common sense.

V. QUALIFICATIONS AND KNOWLEDGE OF ONE OF ORDINARY SKILL IN THE ART

36. Based on my knowledge and expertise, one of ordinary skill in the art at the time of the alleged invention of the '306 Patent would have had a Ph.D. in bioinformatics, molecular genetics, molecular biology or a related field, and at least five years of research in an academic or industry setting, including at least two to three years of research experience in the field of cancer genomics. One of ordinary skill in the art would have had knowledge of DNA sequencing, including next generation sequencing (NGS) and related sequencing methods, and related sample preparation techniques; bioinformatics methods and tools for analyzing sequence reads and mapping sequence reads onto genomes; and methods for identifying genetic variants in a sample.

VI. CLAIM CONSTRUCTION

37. In my opinion, there should not be significant disputes as to the meanings of claim terms of the '306 Patent because the claims are composed of

terms known to one of ordinary skill in the art, who would have understood the meaning of the claim terms in light of the claims, specification, and file history.

38. To the extent Patent Owner modifies its construction for one or more claim term(s) or proposes and/or relies on specific constructions for any claim term in this proceeding, I reserve the right to supplement and modify my opinions.

VII. TECHNICAL BACKGROUND

39. The challenged claims of the '306 Patent are directed to methods of providing a library of cell-free DNA molecules for next-generation sequencing. The challenged claims recite several technical terms and techniques to provide and sequence such libraries, including, but not limited to, the use of cell-free DNA attached with duplex tags, reducing and tracking redundancy based on paired and unpaired sequence reads, and detecting genetic variants in cancer patients. To better understand the appropriate context of the technology at issue, a brief background primer is provided *infra*.

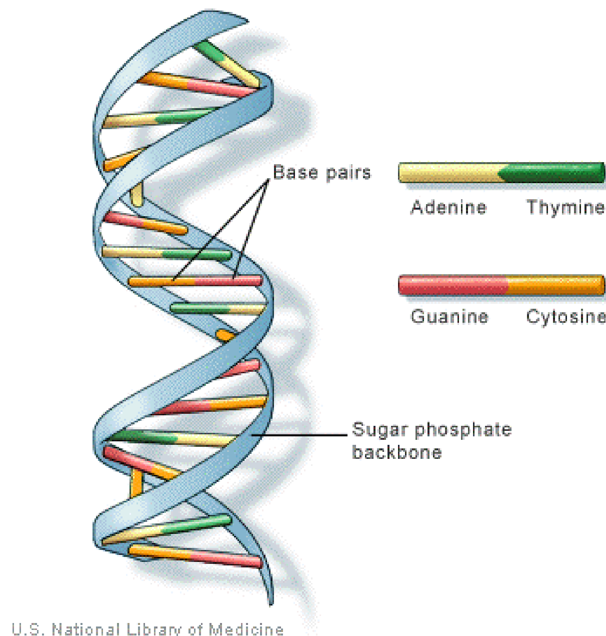
A. DNA Basics

40. DNA is made of four different nucleotides or bases, each represented by a different letter: adenine (A), thymine (T), cytosine (C) and guanine (G). The order of these letters (*i.e.*, the DNA sequence) encodes the information that instructs cells when and how to perform certain functions.

41. The four nucleic acid bases form stable hydrogen bonds according to

specific base-pairing rules. "A" and "T" can form stable hydrogen bonds, and "C" and "G" can form stable hydrogen bonds. Said in another way, "A" only pairs with "T" and does not pair with "G" or "C." Likewise, "C" only pairs with "G" and does not pair with "A" or "T."

42. The formation of base-pairs between nucleotides on the two opposite strands causes the two strands to coil around each other to form a double helix structure, as shown *infra*.



43. According to the base-pairing rules, stretches of nucleotides in one strand can form hydrogen bonds only with a complementary stretch of nucleotides in the other strand. Two single strands of nucleic acids that come together by way of hydrogen bonding between stretches of complementary (or substantially complementary) nucleotides are said to be hybridized to one another. Because of

the specificity of base-pairing, under most environmental conditions, a strand of DNA will have a specific complementary sequence.

B. Human genetics

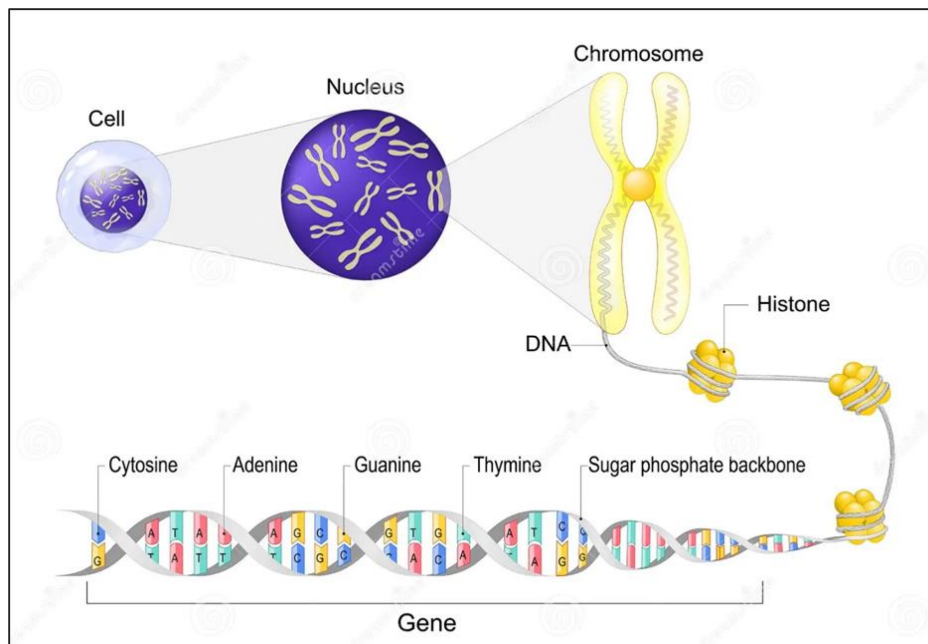
44. DNA is typically contained within the nucleus of a cell and is packed into structures called *chromosomes*. The complete set of DNA found in the nucleus is referred to as a "genome" and is called *genomic DNA* or "gDNA." The *haploid* human genome contains about 3 billion nucleotides that are distributed among 23 chromosomes, and the *diploid* human genome contains about 6 billion nucleotides that are distributed among 46 chromosomes. Said another way, a typical human genome contains 46 chromosomes—44 (22 pairs) autosomes and two (one pair) sex chromosomes (*i.e.*, X and Y).[6] One set of each chromosome pair is inherited from a person's mother and the other from a person's father. In people, almost every cell in the body contains a complete copy of the genome in the nucleus of a cell, organized as chromosomes,[7] as illustrated in the figure *infra*.[8]

---

[6] *See* Ex. 1032 (Buckingham (2007) at 156.

[7] *Id*. Excluded are reproductive cells, called germ cells, which only have one genome copy (*i.e.*, the haploid genome), as well as cells that do not contain a nucleus (*e.g.*, red blood cells), and therefore do not contain genomic DNA.

[8] Image obtained from http://www.cell-anatomy-nucleus-chromosomes-close-up-dna-molecule-histone-sugar-phosphate-backbone-guanine-cytosine-thymine-adenine-247771836.webp/ (last visited July 25, 2025).



C. Cell-free DNA

45. As shown *supra*, DNA is typically contained within the nucleus of a cell and is considered *cellular DNA*. In addition, DNA can exist as *extracellular DNA*, also called *cell-free DNA*, and refers to short fragments of DNA (generally 140-170 bp in length) found in blood.[9] This is because cells typically shed their DNA into the bloodstream through certain biological process associated with cell death (*e.g.*, necrosis or apoptosis) or through active secretion.[10]

D. Genetic Variants & Disease

46. Genetic variants can be associated with cancer.[11] As I explain *infra*,

---

[9] *See* Ex. 1010 (Forshew (2012)) at 1.

[10] *See* Ex. 1014 (Schwarzenbach (2011)).

[11] *See* Ex. 1031 (Thomas (2006)) at Abstract.

such variants may be classified by the type of alteration or cell type in which the variant occurs. I described both types in the section *infra* as well as how genomic variants are identified.

1. *Types of genetic variants*

47. For example, the smallest genetic variants are single nucleotide variants (SNVs). Each SNV reflects a difference in a single nucleotide (or letter) at a particular genomic position. For a given SNV, the DNA letter at that genomic position might be, for example, a "C" in one person but a "T" in another person.[12] Another type of variant is single nucleotide polymorphisms (SNPs), which also refers to a difference in a single nucleotide (or letter) at a particular genomic position. However, to be considered a SNP, a variant is typically present in at least 1% of the human population.[13] As such, an SNV is a more general term that includes both relatively common (such as SNPs) and rare single-nucleotide differences. For simplicity, all single-nucleotide differences can be referred to as SNVs, regardless of their relative frequency.[14]

48. Another class of genetic variants are insertions and deletions, often

---

[12] *See* Ex. 1015 (NHGRI variation fact sheet) at 4.

[13] *Id.*; *see also* Ex. 1032 (Buckingham (2007)) at 156.

[14] *See* Ex. 1015 (NHGRI variation fact sheet) at 4.

called *indels*.[15] A genomic insertion results when one or more nucleotides (usually fewer than 50 base-pairs) are inserted into a DNA sequence *(i.e.*, extra nucleotides found in the genome). A genomic deletion, on the other hand, results when one or more nucleotides (usually fewer than 50 base-pairs) are deleted from a DNA sequence *(i.e.*, missing nucleotides found in the genome). Indels are not as common as SNVs but nonetheless can have a large impact on health and disease. The most common type of indel variants are tandem repeats, which are also called *microsatellites*. These microsatellites, also called *simple sequence repeats* (SSRs) or *simple tandem repeats* (STRs) are tandemly repeated tracts of DNA composed of 1-9 base-pair units.[16] Because of their variability, tandem repeats are also called *variable number of tandem repeats* (VNTRs) and are highly variable among people.[17] One example of a tandem repeat or microsatellite is a (CA)*n* repeat, where *n* can range from a few repeats to over 20 *dinucleotide repeat* units or just *dinucleotide repeats*.[18] Tandem repeats can also contain *trinucleotide repeats* (*e.g.*,

---

[15] *Id*.

[16] *See* Ex. 1033 (Gemayel (2010)) at 446; *see also* Ex. 1034 (Tóth (2000)) at 967.

[17] *See* Ex. 1033 (Gemayel (2010)) at 446.

[18] *See* Ex. 1051 (Sharma (2005)) at 4.

(CAG)*n*), *tetranucleotide repeats* (*e.g.*, (GATA)*n*),[19] and so forth. Microsatellite instability (MSI) occurs in tandem repeats from the accumulation of uncorrected insertion/deletion of repetitive DNA tracts in cancer patients with deficient mismatch repair (MMR) systems[20] and/or aberrant recombination events that can occur between repetitive DNA tracts.[21]

49. Larger indels (*i.e.*, typically > 50 base-pairs) form yet another class of genetic variants called *structural variants*. Tandem repeats that contain more than 50 base-pairs are considered structural variants. Large tandem repeats account for nearly 50% of structural variants in the human genome. When structural variants reflect differences in the total number of base-pairs involved, these are called *copy-number variants* or (CNVs).[22] Several examples of copy number variations, such as large deletions, duplications, and insertions are illustrated *infra*.[23]

---

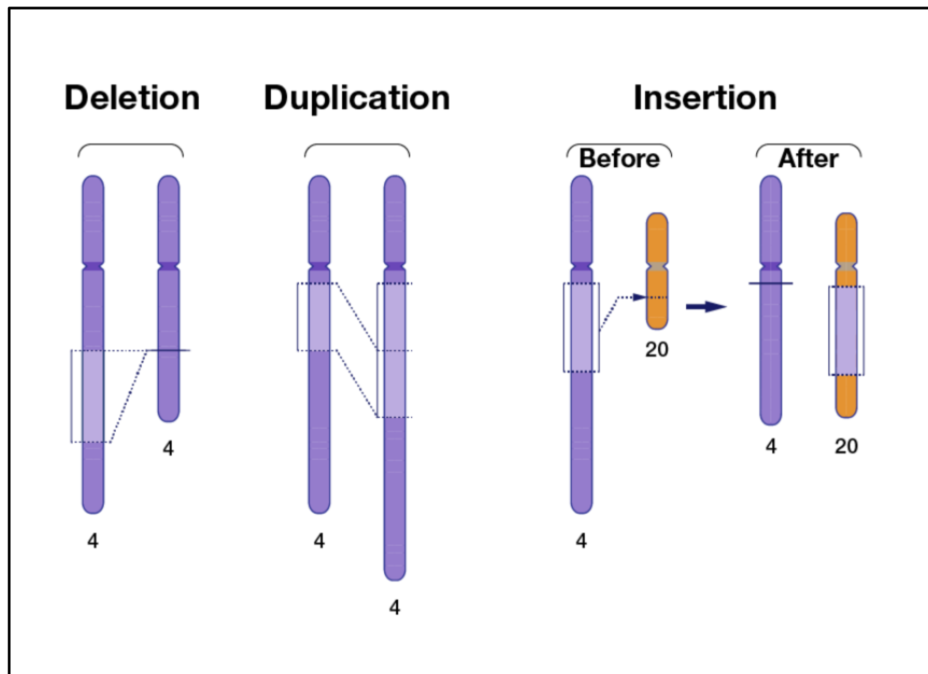
[19] *See* Ex. 1034 (Tóth (2000)) at 970–976; *see also* Ex. 1032 (Buckingham (2007)) at 234.

[20] *See* Ex. 1035 (Laghi (2008)) at Abstract.

[21] *See* Ex. 1036 (Richard (2000)) at Abstract.

[22] *See* Ex. 1037 (Hastings (2009)) at Abstract.

[23] Figure adapted from Ex. 1015 (NHGRI variation fact sheet) at 6.



2. *Cell type: how genomic variants are acquired*

50. Genetic variants or mutations may be inherited, called *germline* variants.[24] Germline variants are first present in reproductive cells (*e.g.*, egg or sperm cells) of a parent, and are then passed from parent to their offspring. Germline transmission typically results in the genetic variant being present in all cells of the offspring.[25]

51. Genetic variants or mutations may also be acquired during a person's lifetime in non-reproductive cells, called *somatic* variants.[26] Somatic variants are

---

[24] *See* Ex. 1038 (Hiatt (2013) at 843.

[25] *See* Ex. 1039 (Cleveland Clinic webpage) at 2–3.

[26] *Id.*; *see* Ex. 1040 (American Cancer Society webpage) at 4.

not heritable and accumulate over a person's lifetime, often due to environmental factors or errors in copying the genome during cell division. Unlike germline mutations, however, somatic variants are not present in all of the cells in a person's body. Rather, somatic variants are only present in the original cell in which the mutation first occurred and then copied and transmitted through cell division to their daughter cells. Some somatic variants can lead to the development of diseases, such as cancer. Unlike germline variants, somatic variants do not typically occur in reproductive cells (*e.g.*, egg or sperm cells) and therefore are not passed on to future generations.[27]

3. *How genetic variants are identified*

52. Comparing the genome sequence derived from a cell of one person against the genome sequence derived from a cell of another person or a reference human genome[28] will generally reveal far more similarities to each other than differences. For example, two people's genomes are generally more than 99% identical to one another because "on average, a single-nucleotide difference exists

---

[27] *Id.*

[28] A "reference" human genome sequence is an established, well-accepted sequence of a human genome. I was part of the effort, called the *Human Genome Project*, that decoded the first human reference genome. *See* Ex. 1041 (IHGSC (2001)); *see also* Ex. 1042 (IHGSC (2004)).

between two people's genomes once every 1,300 nucleotides or so."[29] The differences among human genomes are sometimes called "variants," "variations," "polymorphisms," or "mutations." A person's set of genetic variants is part of what makes them unique[30] and may contribute to an individual's risk for developing diseases, such as cancer.

53. Cataloging those mutations that are associated with cancer and other human disease has long been an area of intense study, dating back decades.[31] Public databases of such genetic variants have long existed.[32] A number of different types of genetic variants have been identified that are associated with cancer, such as SNVs, MSIs, and CNVs, as described *supra*.

E. Sequencing technologies

1. *General principles*

54. DNA sequence technologies are important because they can help diagnose disease by identifying genetic variants. The Human Genome Project begun in October 1990 as a joint effort of the NIH and DOE, with the ultimate goal of determining the complete sequence of the three billion nucleotides that make up the

---

[29] *See* Ex. 1015, (NHGRI variation fact sheet) at 3.

[30] *Id*. at 4.

[31] *See* Ex. 1088 (Fernández-Suárez (2013)).

[32] *See e.g.*, Ex. 1043 (Forbes (2011)); *see also* Ex. 1044 ("GDC data portal").

human genome.[33] I was part of the project that was responsible for sequencing the first human genome at Baylor College of Medicine. It took approximately 13 years to complete the sequence of the first human genome. While the sequencing technologies I described *infra* can produce entire genomes faster and cheaper than the first human genome, doctors or researchers who suspect certain genetic variants that may cause a particular disease will only focus on those parts of the genome that may contain such variants, called *targeted* approaches.

55. DNA sequencing includes several methods and technologies that are used to determine the order of nucleotide bases (that is, A, C, G, and T) in a DNA fragment. Modern methods involve inputting many polynucleotide fragments[34] at once into an automated device called a *sequencer*. After performing a series of chemical and enzymatic reactions to determine the base at each position in the sequences, the sequencer outputs ordered lists of those polynucleotide bases in the form of *sequence reads*. By December 2013, sequencing techniques were routinely and conventionally used for disease detection and other diagnostic methods,[35] and its workflows were well-known to practitioners.

---

[33] I note this is the haploid state, and as diploid organism, the human genome is twice this size.

[34] As described *infra*, not all DNA fragments get sequenced.

[35] *See*, *e.g.*, Ex. 1017 (Metzker (2010)).

2. *Next generation sequencing (NGS)*

56. NGS technologies can produce substantially larger volumes of sequence information (*e.g.*, millions or even billions of sequence reads) at an affordable cost. For example, today a human genome can be sequenced for around $100.[36] Next-generation sequencing (NGS) platforms have allowed for high-throughput sequencing of nucleic acids both on a genome-wide scale and in a targeted, highly multiplexed manner. The most widely used NGS sequencing platform is the Illumina sequencing-by-synthesis (SBS) platform,[37] which has been in use since the 2006. NGS can generate single-end sequence reads from one end of the DNA fragment or paired-end sequence reads from both ends of the DNA fragment.

57. In general, DNA sequencing methods can be characterized as either *first-generation sequencing*, which include Sanger sequencing,[38] or NGS, discussed *infra*. In broad terms, next-generation sequencing technologies are distinguishable from first-generation sequencing technologies in that they produce substantially

---

[36] *See* Ex. 1054 (Preston (2023)). In 2023, the authors wrote "[t]oday, a human genome can be sequenced for $600, with some predicting that the $100 genome is not far behind."

[37] *See, e.g.,* Ex. 1017 (Metzker (2010)) at 34.

[38] *See* Ex. 1045 (Metzker (2005)) at 1768–1770.

larger volumes of sequence information cheaply. I have reviewed many of the platforms using next-generation sequencing technologies.[39] I may rely on figures and information from this article.

58. There have been a number of NGS platforms that have been released commercially since 2005, including those sold by 454 Life Sciences Corporation—the GS20 and GS FLX® instruments,[40] Life Technologies—the SOLiD® instrument,[41] and Pacific Biosciences—the single-molecule real-time or SMRT® instrument.[42] Other NGS systems that have been commercialized include the Ion Torrent—the Personal Genome Machine ("PGM®")[43] and Ion Proton® instruments and Oxford Nanopore—the MinION sequencer.[44] The Genome Analyzer ("GA") sequencer, called the GA1, was developed by Solexa and released in late 2006. The GA II sequencer was developed after the acquisition of Solexa by Illumina in 2007. Since 2009, Illumina has released new instruments including multiple versions of HiSeq®, MiSeq®, NextSeq® and NovaSeq® systems.

---

[39] *See, e.g.,* Ex. 1017 (Metzker (2010)).

[40] *Id.* at 39, Figures 1a, 3c, 3d.

[41] *Id.* at 36–39 at Figures 1a, 3a, 3b.

[42] *Id.* at 39 at Figures 1e, 4.

[43] *See* Ex. 1046 (Rothberg (2011)).

[44] *See* Ex. 1047 (Jain (2015)).

3. *Only a subset of NGS sequence reads is generated*

59. Many NGS workflow schematics suggest that *all the sample materials* that are prepared for NGS analysis actually get loaded onto the sequencer. In fact, *only a limited portion* of the NGS fragment molecules make it onto the sequencer, leaving behind many DNA fragments. Relevant for this matter, many DNA fragments representing *only one strand of the double-stranded molecule get loaded onto the sequencer*. The primary reason for the sample reduction is the clonally amplification process itself, such as the bridge amplification step performed on the Illumina system. This is because single DNA fragment molecules must be sufficiently and spatially distributed across the flow cell to yield useful data. The introduction of too few fragments on the flow cell results in little useful data while adding too many fragments on the flow cell results in copies of DNA fragments crowding one another, yielding noisy (and not useful) sequence data.

60. For example, Bentley (2008)[45] was the first Illumina paper to describe their SBS method using the GA1 analyzer.[46] Bentley (2008) explained to its users that NGS libraries *must be diluted* to a concentration of 10 nanomolar ("nM") prior to bridge amplification on the flow cell. "***Samples were diluted to a concentration***

---

[45] *See* Ex. 1055 (Bentley (2008)).

[46] *Id*. at 53.

of 10 nM in 10 mM Tris pH 8.5 and 0.1% Tween 20 ***prior to cluster formation*** (see below)."[47] Quail (2009)[48] also explained that accurate quantification of the library *is vital*, as if the concentration is too low or too high, data yield will be reduced.

> The number of clusters per lane of a flowcell is governed by the concentration of library that is added. ***Accurate quantification is vital***, because ***too low a cluster density reduces the yield of data***, and therefore increases the per-base cost of sequencing, whereas ***too high a cluster density results in a reduced yield of data*** due to cluster overlap.[49]

For duplex sequencing efforts, the overall result is a population of sequence reads that are *unpaired* (*i.e.*, only one of the two strands on a double-strand DNA molecule was sequenced). As the dilution can affect which of the DNA fragments get loaded onto the sequencer, there will also be a population of sequence reads that are *paired* (*i.e.*, both strands on a double-strand DNA molecule were sequenced).

F. Amplification and target enrichment of multiple loci

1. *Polymerase chain reaction (PCR) amplification*

61. Sequencing is generally not performed on the DNA fragments as they are originally collected from a sample. Prior to sequencing, the fragments are generally modified and/or amplified to create multiple copies of each DNA

---

[47] *See* Ex. 1055 (Bentley (2008), Supplementary Information) at 2 (emphases added).

[48] *See* Ex. 1087 (Quail (2009).

[49] *Id*. at 18.2.3 (emphases added).

fragment. This is typically necessary to generate sufficient copies for the sequencing process to produce usable data.

62. A universal technique for replicating or amplifying nucleic acids in a laboratory setting is the *polymerase chain reaction* (PCR). This method was invented by Dr. Kary Mullis in the 1980s.[50] PCR is used to specifically amplify a small amount of nucleic acid molecules, generating thousands to millions of copies of the target nucleic acid.

63. Generally speaking, PCR involves repeated denaturation and replication of DNA sequences *in vitro* (*i.e.*, outside of an organism). The enzyme *DNA polymerase* copies or synthesizes the complementary strand from a single-stranded template. For this enzymatic reaction to occur, a partially double-stranded section of DNA is required. Typically, a primer hybridizes to a complementary region of a single-stranded template. Dr. Mullis and colleagues used short oligonucleotide primers complementary to the 3'-ends of the sequence of interest. Because of base-pairing rules, primer sequences can be designed to target specific nucleic acid sequences. DNA polymerase initiates synthesis of the nascent strand from the hybridized primers in a 5'-to-3' direction to create double-stranded DNA. Two primers are used for each template strand in PCR that define the targeted

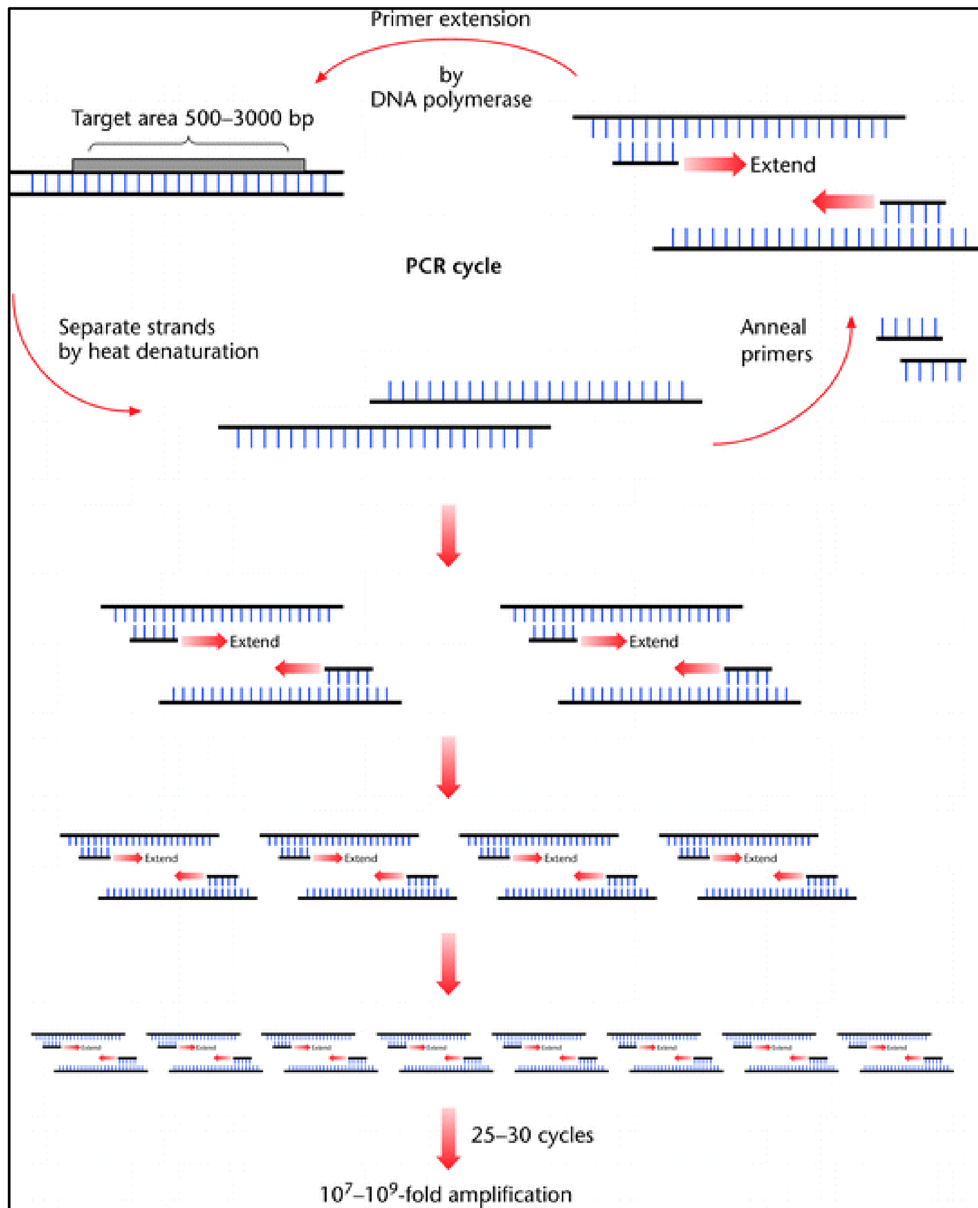
[50] *See* Ex. 1019 (Metzker (2006)) at 380.

sequence of interest: a *forward* primer will hybridize to the 3'-end of a single-stranded template DNA. DNA polymerase binds to this partially double-stranded complex and extends the primer by copying the sequence of the single-stranded template in the 5'-to-3' direction. Primer extension will build a complementary strand of DNA.[51] Also in the same cycle, a *reverse primer* will hybridize to the 3'-end of the other single stranded template DNA. Similarly, DNA polymerase binds to this partially double-stranded complex and extends the primer by copying the sequence of the single-stranded template in the 5'-to-3' direction. Primer extension will build another complementary strand of DNA. The general concept of PCR is illustrated by the figure *infra*.[52]

---

[51] *Id*.

[52] *Id*. at Figure 1.



64. Amplification can be *universal*—amplifying all molecules present in a sample equally—or *targeted*—amplifying just those that are of interest. The latter approach is a process referred to as *target enrichment*. By using specially designed primers to preferentially amplify sequences of interest over other background sequences that are present, downstream sequencing resources can be used efficiently

and cost effectively. After this process, a sample of interest is said to be enriched. Several target enrichment methods have been described that target multiple loci in a single reaction tube to increase the proportion of specific target nucleic acid sequences in a biological sample,[53] as described *infra*.

2. *Multiplex PCR*

65. Multiplex PCR refers to amplifying multiple different target sequences in a target DNA using different target-specific primer pairs in the same reaction tube or volume.[54] Multiple different targets can also be amplified in separate PCR reactions, called single-plex or uniplex reactions, though this is generally more time-consuming and labor-intensive.[55]

66. In targeted multiplex PCR, target-specific primers operate in the same manner as in single-plex PCR, whereby primer pairs for each target region undergo the PCR cycle, as illustrated in the figure *supra*. Rather than having only one set of primers (a forward primer and a reverse primer) in a reaction as in single-plex PCR, however, multiplex PCR requires at least two or more sets of target-specific primer pairs, and can include dozens of different primer pairs in the same reaction.[56] The

---

[53] *See* Ex. 1020 (Mamanova (2010)) at 111.

[54] *Id.* at Figure 1 legend; *see also* Ex. 1049 (Turner (2009)) at 266.

[55] *See* Ex. 1050 (Edwards (1994)) at S66.

[56] *See* Ex. 1049 (Turner (2009)) at 269.

mixture of many target-specific primer pairs with different sequences allows targeting of many sequences located within the template DNA, and results in amplification of a plurality of specific targets.[57]

67. A challenge with PCR, particularly multiplex PCR, is the formation of unwanted amplification products called, *primer dimers*.[58] Primer dimers form when target-specific primers hybridize to one another rather than the target sequence and typically result because the primers in the reaction are in relatively high concentrations compared to the template DNA.[59] The presence of multiple primers in multiplex PCR tends to increase the propensity of primer dimer formation, which has the potential to obscure downstream sequencing by increasing the amount of nonspecific background noise.[60]

3. *Hybrid capture*

68. Several hybridization-based capture methods have been developed to target genomic sequences of interest, most of which can be employed after nucleic acid extraction and library preparation.[61] One such solution-based hybridization

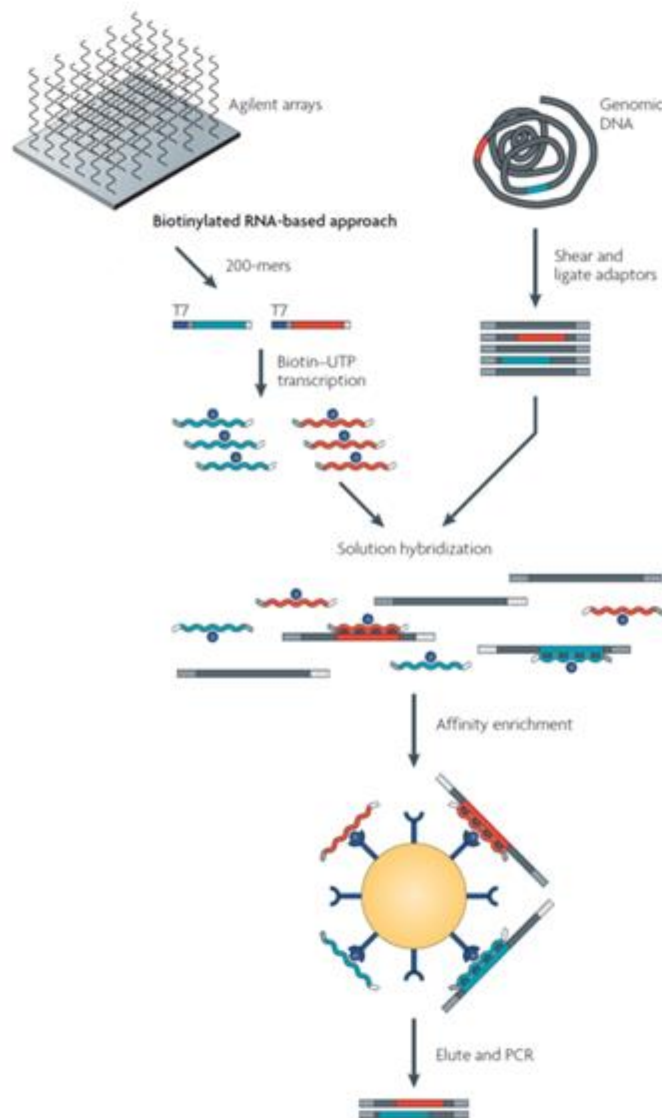
---

[57] *See* Ex. 1019 (Metzker (2006)) at 384.
[58] *See* Ex. 1050 (Edwards (1994)) at S69.
[59] *Id*.
[60] *Id*.
[61] *See* Ex. 1017 (Metzker 2010)) at 41, Figure 5.

method employs the use of RNA "baits." Generally, DNA fragments are first ligated with adaptor sequences to create adaptor-ligated libraries, as described *infra*. These libraries are denatured, for example, by exposure to heat. The denatured adaptor-ligated products are then hybridized with single-stranded, biotinylated RNA baits that are specific to the region(s) of interest. The resulting double-stranded, hybridizes duplexes are capture by streptavidin-coated magnetic beads. A magnetic field is applied, capturing the bead-DNA complexes. This is illustrated in the figure *infra*.[62]

---

[62] *Id*. at Figure 5c.



69. Adapter-ligated DNA fragments that are not targeted, and therefore not captured by RNA baits, are then washed away. The remaining enriched adaptor-ligated DNA is eluted and PCR amplified to create a capture hybrid enriched library that can be sequenced by NGS technologies. The combination of genome enrichment methods, such as hybrid capture, and next generation sequencing technologies have facilitated an ". . . increase in the efficiency and resolution of

detection of each of the principle types of somatic cancer genome alterations, including nucleotide substitutions, small insertions and deletions, copy number alterations, chromosomal rearrangements and microbial infections."[63]

G. Library preparation, Including Tagging With Barcodes

70. Before NGS, the nucleotide fragments of interest are typically combined with relatively short pieces of DNA called *adaptors*[64] that serve a variety of functions. This modification step is referred to as *tagging*. These adaptors can added through the amplification process discussed above (*e.g.*, by PCR)[65] or alternatively through direct attachment (*e.g.*, by a process called *ligation*).[66]

71. In Illumina sequencing, these tagged adaptors always included known primer sequences necessary for the sequencing process itself. The adaptors also frequently included unique or non-unique identifier tags which could, for example, serve to distinguish original DNA fragments and their resulting sequence reads from one another. Such identifiers, which had long been used in other molecular biology

---

[63] *See* Ex. 1053 (Meyerson (2010)) at Abstract.

[64] "Adaptors" have also been spelled as "adapters." Both spellings have been used in many of the documents cited in this Declaration, and their descriptions therein should be considered to have identical functions.

[65] *See* Ex. 1010 (Forshew (2012)), Supplementary Methods at 3.

[66] *See* Ex. 1055 (Bentley (2008)) at Figure 1; *see also* E. 1056 (Kircher (2012)) at 2.

techniques (such as array hybridization) before their use in NGS, are typically referred to as barcodes.[67]

72. Barcodes can be unique or non-unique. In the former case, there are more barcodes than fragments of interest—consider using 1000 barcodes to label 500 DNA fragment of interest. In the latter case, there are less barcodes than fragments. Unique tagging can still be achieved using non-unique barcodes, however, by considering the barcodes together with some portion of the sample fragments themselves (*e.g.*, a few nucleotides).[68]

73. Tagging with adaptors can occur on single-stranded DNA (at one or both ends), or in the case of double-stranded DNA fragments, both strands (again at one or both ends). This latter approach is called "duplex tagging,"[69] and was also well known by December 28, 2013.

H. Sequence reads

74. The output of sequencing using NGS methods is typically a large dataset of DNA sequences. These sequences are commonly referred to as *sequence*

---

[67] *See* Ex. 1057 (Pierce (2006)).

[68] *See* Ex. 1073 ('389 PCT Appl. (Schmitt)) at [0038]. ("***A hybrid method*** using a combination of sheared ends and a shorter n-mer tag (such as 1 or 2 or 3 or 4 or more degenerate or semi-degenerate bases) in the adaptor ***may also serve as unique molecular identifiers***.") (emphases added).

[69] *See* Ex. 1056 Kircher (2012)); *see also* Ex. 1058 (Schmitt (2012)).

reads. Each sequence read represents the order of the nucleotide bases that was detected by the sequencer. Typically, the length of a sequence read is governed by the type of NGS platform used. For example, by 2012, Illumina sequencers produced 150 base reads, Ion Torrent sequencers produced 400 base reads, and PacBio sequencers produced ~1,000 base reads.[70] In addition to determining the order of the sequence, the sequence reads also contain quality information of the base call, called a *quality score*. These Q-scores are expressed as probability errors. For example, a base call associated with a Q-score of 20 means that call has a probability error of 1% or 0.01.[71] The data are typically stored in standardized file formats (*e.g.*, FASTQ file)[72] and can be further analyzed to determine the order of nucleotides in the original sample, identify variants, and perform other downstream analyses, as described *infra*.

I. Mapping/Alignment

75. After sequence reads are generated, they are usually compared to a reference DNA sequence to know where in the genome the sample DNA fragments came from. This bioinformatic process is called *mapping* or *read alignment*. There

---

[70] *See* Ex. 1059 (Glenn (2011)) at Table 2.

[71] *See* Ex. 1061 (CASAVA 1.8 User Guide) at 146. Likewise, a Q-score of 30 means a base call has a probability error of 0.1% or 0.001.

[72] *See* Ex. 1060 (Cock (2010)).

are multiple alignment tools that have been developed to map sequence reads to a reference genome. For example, such alignment tools include programs called ELAND,[73] MAQ,[74] BWA,[75] Bowtie,[76] and CASAVA,[77] just to name a few. Once sequence reads are mapped reads to the reference sequences, they are formatted into a common alignment format, called Sequence Alignment/Map (SAM) format, which provides a well-defined interface between alignment and downstream analysis.[78] A companion format, called a Binary Alignment/Map (BAM) format or BAM file was developed that represents the same information as a SAM file, but in a compressed memory format.[79]

76. Processing such an enormous volume of NGS sequence data require intensive computer efforts. For example, mapping tens to hundreds of millions of sequence reads using alignment algorithms to a reference sequence to create BAM files requires a significant computer infrastructure utilizing multiple centralized

---

[73] *See* Ex. 1055 (Bentley (2008)) at 54; *see also* Ex. 1061 (CASAVA 1.8 User Guide) at 6.

[74] *See* Ex. 1062 (Li (2008)).

[75] *See* Ex. 1063 (Li & Durbin (2009)); *see also* Ex. 1064 (Li & Durbin (2010)).

[76] *See* Ex. 1065 (Langmead (2009)).

[77] *See* Ex. 1061 (CASAVA 1.8 User Guide).

[78] *See* Ex. 1066 (Li (2009)).

[79] *Id*.

processing units (CPUs).[80] In the case of human DNA, such reference sequences have long been made available by the research community as public resources.[81] In 2010, the Broad Institute created the Genome Analysis Toolkit (GATK) that provided a structured programming framework to bridge the development gap that existed between NGS read production and genetic variant analysis, the framework of which was based on Google's programming philosophy of MapReduce.[82] Development of the GATK engine has been an indispensable platform that is constantly being improved and optimized, even today for "correctness, stability, and CPU and memory efficiency"[83] for genome analysis.

77. Mapping sequence reads to a reference sequence facilitates the identification of genetic variants, a process referred to as *variant calling* because a DNA fragment can be identified as containing a genetic variant when its sequence is different from the reference at one or more nucleotide positions. There were also a

---

[80] *See* Ex. 1067 (DePristo (2011)) at 491 ("Mapping reads to the reference genome is a first critical computational challenge . . .") (citations omitted).

[81] *See* Ex. 1041 (IHGSC (2001)); *see also* Ex. 1042 (IHGSC (2004)).

[82] *See* Ex. 1068 (McKenna (2010)) at 1297.

[83] *Id*; *see also* Ex. 1069 (GATK Updates).

variety of variants callers, such as VarScan,[84] VarScan2,[85] MuTect,[86] and CASAVA,[87] just to name a few. For example, these variant callers could analyze a BAM file of a sequence read to identify an "A" at a position where the reference sequence identifies a "T," which would characterize this genetic variant as an SNV. The output of variant callers is provided in another standardized format, called the *variant call format* (VCF).[88]

78. However, sequence read data created by NGS was not error-free,[89] and in fact, some high-quality base calls can be errors. In some cases, mapping a single sequence read alone (or even mapping a handful of reads) is generally not sufficient to know whether a difference between a given sequence read and a reference sequence is, in fact, a true mutation in the original sample. That is, in the example *supra*, the "A" in the sequence read could not reflect a genetic variant at all but rather

---

[84] *See* Ex. 1070 (Koboldt (2009)).

[85] *See* Ex. 1071 (Koboldt (2012)).

[86] *See* Ex. 1072 (Cibulskis (2013)).

[87] *See* Ex. 1061 (CASAVA 1.8 User Guide).

[88] *See* Ex. 1089 (Danecek (2011)).

[89] Even at 99.9% accuracy (called Q30 by practitioners), there will be an error—a base called incorrectly—at 1 out of every 1000 bases. When billions of bases are being called in a single sequencing run, there will be a significant number of errors spread across the resulting sequence reads; *see also* Ex. 1017 (Metzker (2010)) at 34; Ex. 1048 (Kinde (2011)); Ex. 1058 (Schmitt (2012)).

could represent a *sequencing error*.[90] Sequence redundancy, using multiple sequence reads in an assembly or alignment is well established to reduce errors in individual sequence reads by generating a *consensus sequence*.[91]

J. Generating consensus sequences

79. Early on, practitioners recognized the need for high quality sequence reads to accurately call genetic variants, particularly those that are rare and thus not expected to be widely present in subject DNA. Thus, while efforts continue to improve the sequencing process itself to increase its fidelity, researchers also developed other techniques to increase the accuracy of detecting rare genetic variants.

80. One such technique that relies on the barcodes discussed earlier is known as generating consensus sequences.[92] It is undisputed that generating consensus sequences was well-known in the art by December 28, 2013.[93] At a high level, generating consensus sequences first involves bioinformatically "grouping" together similarly-tagged sequence reads (those having the same barcode identifier)

---

[90] *See* Ex. 1048 (Kinde (2011)); Ex. 1058 (Schmitt (2012)).

[91] *See*, *e.g.*, Ex. 1090 (Schatz (2010)) at 1167 ("These assemblers also attempt to correct sequencing errors by using overlapping reads to confirm each other.").

[92] *See* Ex. 1048 (Kinde (2011)); Ex. 1058 (Schmitt (2012)).

[93] *See* Ex. 1090 (Schatz (2010)).

into sets of reads called "families." For each family, those grouped sequence reads are then "collapsed" into a single, accepted consensus sequence based on the sequence that appears most frequently (*e.g.*, a majority rule). This consensus sequence is expected to be the most accurate because sequence reads with sequencing errors at a given position will generally be present at a very low frequency.[94]

81. For example, while one sequence in a family may have an incorrect base at a particular position due to an error, it will be expected to be out-represented in that family by other sequences without an error at that position.

82. Generating consensus sequences thus results in increased practitioner confidence in the variant identification, leading to better and more actionable data. It has wide application and can be used to detect all manner of variants, including those discussed above, with improved accuracy.

K. Copy number variant analysis

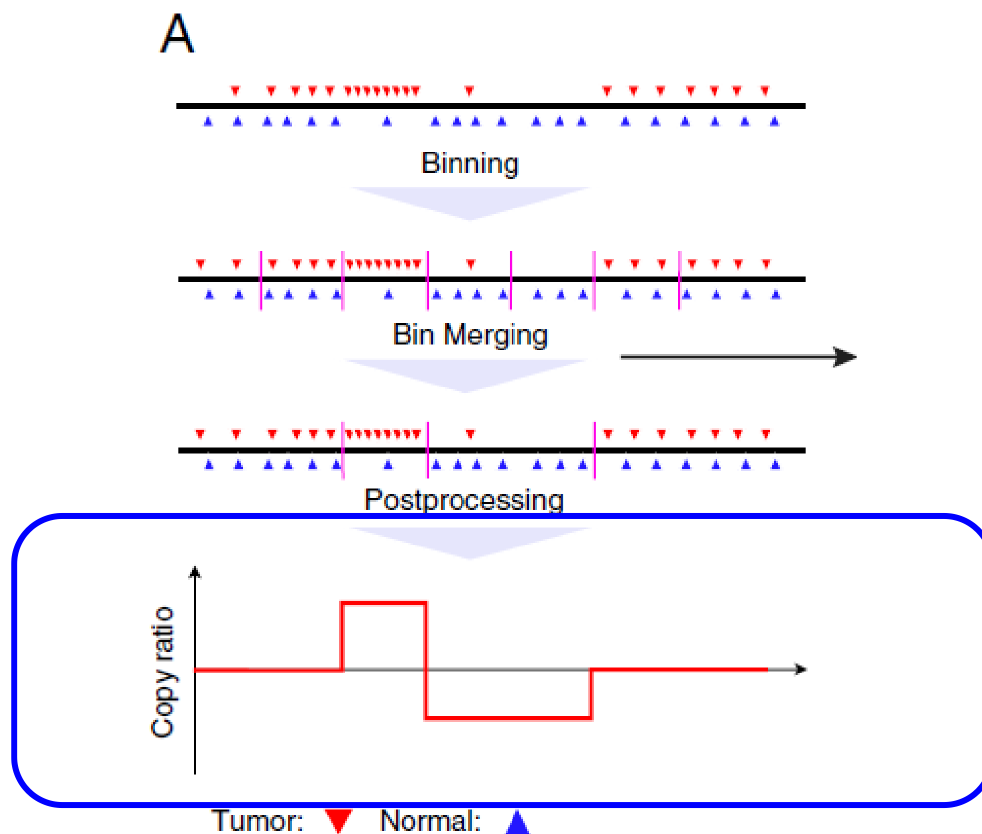
83. As described *supra*, large insertion/deletion regions are known as structural variants or *copy number variants* (CNVs).[95] There are several approaches to analyzing CNVs from NGS sequence read data. One approach uses a *binning*

---

[94] *See* Ex. 1048 (Kinde (2011)); Ex. 1058 (Schmitt (2012)).

[95] *See* Ex. 1091 (Cooper (2007)).

approach that analyzes sequence reads mapped to a reference genome in discrete windows along the genome. Here, the mapped reads derived from a normal and tumor sample are analyzed by position and frequency, which are then binned into discrete window (red vertical lines shown *infra*) and quantified to determine a copy number ratio of the region of interest, as illustrated *infra*.[96] The



horizontal line with the arrow inside the blue box represents the normal copy number of a particular region of the genome. In other words, it is the normalized baseline

[96] *See* Ex. 1086 (Xi (2011)) at Figure 1.

from which CNV gains (*i.e.*, regions above the normal) and CNV losses (*i.e.*, regions below the normal) can be observed.

VIII. THE '306 PATENT

A. Overview And Independent Claims

84. The '306 Patent issued on October 19, 2021 from U.S. Patent Application No. 16/945,124, which was filed July 31, 2020. The '306 Patent claims priority to provisional application no. 61/921,456, which was filed on December 28, 2013. For purposes of my analysis and opinions in this declaration I have assumed December 28, 2013 to be the priority date for the '306 Patent.

85. The '306 Patent contains 29 claims, two of which are independent (Claims 1 and 17). Independent Claim 1 of the '306 Patent recites the following:

> 1[pre]. A method, comprising:
>
> [1a] (a) providing a population of cell-free deoxyribonucleic acid (cfDNA) molecules having first and second complementary strands;
>
> [1b1] (b) tagging a plurality of the cfDNA molecules in the population with duplex tags comprising molecular barcodes to produce tagged parent polynucleotides, wherein the duplex tags are attached to both ends of a molecule of the plurality of the cfDNA molecules,
>
> [1b2] wherein the plurality of the cfDNA molecules are tagged with n different combinations of molecular barcodes, wherein n is at least 2 and no more than 100,000*z, wherein z is a mean of an expected number of duplicate molecules in the population of cfDNA molecules that map to identical start and stop positions on a reference sequence;
>
> [1c] (c) amplifying a plurality of the tagged parent polynucleotides to produce amplified progeny polynucleotides;

[1d] (d) sequencing at least a subset of the amplified progeny polynucleotides to produce a set of sequence reads; and

[1e] (e) reducing or tracking redundancy of a plurality of sequence reads from the set of sequence reads using at least sequencing information from the molecular barcodes of the duplex tags to determine distinct cfDNA molecules from among the tagged parent polynucleotides, wherein the distinct cfDNA molecules are determined based on (i) paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides, or (ii) unpaired reads corresponding to sequence reads generated from a first tagged strand having no second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides,

[1f] wherein reducing or tracking the redundancy of the plurality of sequence reads comprises mapping at least a subset of the plurality of sequence reads to the reference sequence.

86. Independent Claim 17 of the '306 Patent recites the following:

17[pre] A method, comprising:

[17a] (a) tagging a population of double-stranded cell-free deoxyribonucleic acid (cfDNA) molecules obtained or derived from a sample of a subject with a set of tags comprising molecular barcodes to produce tagged parent polynucleotides;

[17b] (b) amplifying a plurality of the tagged parent polynucleotides to produce amplified progeny polynucleotides;

[17c] (c) sequencing at least a subset of the amplified progeny polynucleotides to produce a set of sequence reads; and

[17d] (d) sorting a plurality of sequence reads from the set of sequence reads into (i) families comprising paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand derived from double-stranded cfDNA molecules from among the tagged parent polynucleotides, and (ii) families comprising unpaired reads corresponding to sequence reads

generated from a first tagged strand having no second tagged complementary strand derived from double-stranded cfDNA molecules from among the tagged parent polynucleotides.

87. The ’306 Patent acknowledges that “many methods have been developed for accurate copy number variation estimation, especially for heterogeneous genomic samples, such as tumor-derived gDNA or for cfDNA for many applications” and “these methods include sample preparation whereby the original nucleic acids are converted into a sequenceable library, followed by massively parallel sequencing, and finally bioinformatics to estimate copy number variation at one or more loci.”[97] According to the ’306 Patent, “[a]lthough many of these methods are able to reduce or combat the errors introduced by the sample preparation and sequencing processes for all molecules that are converted and sequenced, these methods are not able to infer the counts of molecules that were converted but not sequenced.”[98] Thus, the ’306 Patent indicates that it is directed to providing a method by which “[t]he number of unseen molecules can be estimated based on the number of Pairs and Singlets detected.”[99]

88. Although the ’306 Patent professes to address this purported need in the

---

[97] *See* Ex. 1001 (’306 Patent) at 1:46–55.

[98] *Id.* at 1:59–63.

[99] *Id.* at 2:17–18.

art for a way to estimate "the number of unseen molecules," the claims of the '306 Patent are not directed to such a method. Instead, the claims cover sequencing methods that were well-known to one of ordinary skill in the art before December 28, 2013. Specifically, the claims recite methods that include tagging polynucleotides with molecular barcodes,[100] amplifying the tagged parent polynucleotides to produce amplified progeny polynucleotides, sequencing the amplified progeny polynucleotides, and either (*i*) reducing redundancy by grouping sequence reads into families of paired or unpaired reads (in the case of independent Claim 1 of the '306 Patent) or (*ii*) sorting reads into families comprising paired reads and families comprising unpaired reads (in the case of independent Claim 17 of the '306 Patent). As such, these claims recite NGS library and sequencing methods that produce conventional consensus sequences that were already known in the art.

89. Indeed, the '306 Patent generally concedes that the techniques in its claimed methods were readily available, as summarized *infra*.

90. **Tagging with barcodes.** The '306 Patent states that "[t]agging disclosed herein can be performed using any method"[101] and refers to prior art

---

[100] *See* Ex. 1001 ('306 Patent) at 23:53–55 ("A molecular barcode (also 'barcode' or 'identifier' herein) sequence is a nucleotide sequence that distinguishes one polynucleotide from another.").

[101] *Id.* at 17:35–36.

references that disclose tagging with either unique or non-unique barcodes.[102] Tagging with adapters that include molecular barcodes (also known as unique identifiers) was well-known in the art by December 28, 2013.[103]

91. As was known in the art, the ’306 Patent states that the number of barcodes used can vary significantly depending on the needs, including from “at least the same as the estimated number of molecules in the sample” to “at least two, three, four, five, six, seven, eight, nine, ten, one hundred or one thousand times as many as the estimated number of molecules in the sample.”[104]

92. **Amplifying tagged parent polynucleotides.** The ’306 Patent states that amplification is done by routine methods, including PCR.[105] PCR amplification was well-known in the art by December 28, 2013.[106]

---

[102] *Id.*, 22:19–25 (citing, *e.g.*, Ex. 1007 (’078 Patent Appl. (Fodor)).

[103] *See supra* at § VII.G.

[104] *See* Ex. 1001 (’306 Patent) at 18:26–19:21; *see also id.* at 22:34–66 (“Identifiers or molecular barcodes used herein can comprise any types of oligonucleotides. . . . For example, the nucleotide sequence can comprise 1 (if using a non-natural nucleotide), 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides.”).

[105] *Id.* at 29:66–30:2 (“Nucleic acid amplification techniques can be used with the assays described herein. . . . For example, amplification may comprise PCR-based amplification.”); *see also id.* at 30:5–17 (listing known polymerases used for amplification).

[106] *See supra* at § VII.F.1.

93. **Sequencing.** The ’306 Patent states that “[t]he sequencing method can be massively parallel sequencing” using any of several conventional methods listed (including Illumina sequencing) “or any other sequencing methods known in the art.”[107] Sequencing methods were well-known in the art by December 28, 2013.[108]

94. **Reducing redundancy or sorting reads.** The ’306 Patent states that “[s]equencing of amplified polynucleotides can produce reads of the several amplification products from the same original polynucleotide, referred to as ‘redundant reads.’”[109] The ’306 Patent describes multiple known methods for identifying redundant reads, including “based on their distinct barcode,” “based on sequences at the beginning and end of a read, optionally in combination with the length of the read,” or “combining information from a barcode with start/stop sequence and/or polynucleotide length.”[110] Such methods were-known in the art in the art by December 28, 2013.[111]

95. The ’306 Patent also describes that “[c]onsensus sequences can be generated from families of sequence reads by any of a variety of methods” to reduce

---

[107] *See* Ex. 1001 (’306 Patent) at 30:51–67.

[108] *See supra* at § VII.E.

[109] *See* Ex. 1001 (’306 Patent) at 31:34–38.

[110] *Id.* at 31:38–58.

[111] *See supra* at § VII.G.

or track redundancy, and identifies several known methods.[112] Methods for generating consensus sequences by grouping single strand reads or duplex-strand reads were well-known in the art by December 28, 2013.[113]

B. ’306 Patent Prosecution History

96. The ’306 Patent issued on October 19, 2021 from U.S. Patent Application No. 16/945,124, which was filed on July 31, 2020. Patent App. No. 16/945,124 is a continuation of App. No. 16/601,168, which is a continuation of Patent App. No. 15/892,178, which is a continuation of Patent App. No. 14/861,989, which is a continuation of PCT/US2014/072383 filed on December 24, 2014. The ’306 Patent also claims priority to Provisional App. No. 61/948,509 filed on March 5, 2014, and Provisional App. No. 61/921,456 filed on December 28, 2013.

97. On September 9, 2020, the Examiner issued a non-final rejection of pending claims 1-30, which included rejections of certain claims as indefinite, and on the grounds of nonstatutory double patenting.[114] The applicants responded to the first office action by making some amendments to the claims and by filing a terminal disclaimer.[115]

---

[112] *See* Ex. 1001 (’306 Patent) at 31:30–32:29.

[113] *See supra* at § VII.J.

[114] *See* Ex. 1002 (’306 Patent File History) at pp. 360–367.

[115] *Id*. at pp. 370–379, 529–530.

98. On December 21, 2020, the Examiner issued a final rejection of the pending claims, finding that pending claims 1-2, 9, 11-12, 14-15, 18-19, 21, 23 and 25-29 were anticipated by U.S. Patent No. 10,752,951 to Salk ("Salk"), and claims 3-6, 8, 10, 13, 16-17, 20, 22, 24 and 30 were obvious in view of Salk because "they clearly merely relate to routine optimization of known-important reaction parameters."[116]

99. Following an Examiner Interview on February 25, 2021,[117] the applicants again amended the claims, including the following amendment to pending claim 1:[118]

---

[116] *Id*. at pp. 542–547.

[117] *Id*. at p. 550.

[118] *Id*. at pp. 551–556.

1. (Currently amended): A method, comprising:

(a) providing a population of cell-free deoxyribonucleic acid (cfDNA) molecules having first and second complementary strands;

(b) tagging a plurality of the cfDNA molecules in the population with duplex tags comprising molecular barcodes from a set of molecular barcodes to produce tagged parent polynucleotides, wherein the duplex tags are attached to both ends of a molecule of the plurality of the cfDNA molecules;

(c) amplifying a plurality of the tagged parent polynucleotides to produce amplified progeny polynucleotides;

(d) sequencing at least a subset of the amplified progeny polynucleotides to produce a set of sequence reads; and

(e) reducing ~~and/~~or tracking redundancy ~~in~~ of a plurality of sequence reads from the set of sequence reads using at least sequencing information from the molecular barcodes of the duplex tags to determine distinct cfDNA molecules from among the tagged parent polynucleotides, wherein the distinct cfDNA molecules are determined based on (i) paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides, ~~and/~~or (ii) unpaired reads corresponding to sequence reads generated from a first tagged strand having no second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises mapping at least a subset of the plurality of sequence reads to a reference sequence.

The applicants also presented several arguments to distinguish Salk from the pending claims, including that Salk purportedly did not disclose "a method directed to the analysis of cell-free deoxyribonucleic acid (cfDNA) molecules)" or the added limitation of claim 1 that "wherein reducing or tracking the redundancy of the plurality of sequence reads comprises mapping at least a subset of the plurality of

sequence reads to a reference sequence."[119]

100. On May 18, 2021, the Examiner maintained the rejection of the claims as anticipated and obvious in view of Salk.[120] The Examiner contended that, contrary to the applicants arguments, Salk disclosed "grouping reads into families is disclosed in the paragraph bridging columns 26 and 27, and mapping reads to a reference sequence is disclosed in column 28 at line 45 as well as in claims 2-4, 8-9, 13-14, and 26."[121] With respect to the claims that were rejected as obvious, the Examiner again found that "they clearly merely relate to routine optimization of known-important reaction parameters, which as well established in U.S. patent practice does not support unobviousness."[122] The Examiner further noted that the applicants arguments had been considered but were not persuasive, including because "as one of ordinary skill in the art is aware, 'circulating' nucleic acid molecules are the same as 'cell-free' or 'cf' nucleic acid molecules."[123]

101. On August 18, 2021, applicants amended claim 1 to add that the "molecules are tagged with n different combinations of molecular barcodes, wherein

---

[119] *Id*. at pp. 656-660.

[120] *Id*. at pp. 675–679.

[121] *Id*. at p. 677.

[122] *Id*. at p. 678.

[123] *Id*.

n is at least 2 and no more than 100,000*z, wherein z is a mean of an expected number of duplicate molecules in the population of cfDNA molecules that map to identical start and stop positions on a reference sequence."[124] Applicants argued that this limitation was not found in Salk.[125]

102. On September 2, 2021, the Examiner allowed the claims without explanation.[126]

IX. THE PRIOR ART REFERENCES

A. PCT Publication No. WO 2013/123442 A1 ("PCT '442 Appl. (Bielas)")

103. PCT '442 Appl. (Bielas) is a patent application filed on February 15, 2013 and published on August 22, 2013. I understand that it is prior art under at least §102(a)(1).

104. PCT '442 Appl. (Bielas) discloses methods for accurately detecting mutations in DNA sequences by tagging double-stranded nucleic acid molecules with duplex molecular barcode tags, called "cyphers" or "barcodes."[127] In PCT '442 Appl. (Bielas), "'random cypher' or 'cypher' or 'barcode' or 'identifier tag' and

---

[124] *Id*. at pp. 684–689.

[125] *Id*. at p. 694.

[126] *Id*. at pp. 737–742.

[127] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 7:23–27; 30:12–13.

variants thereof are used interchangeably and refer to a nucleic acid molecule having a length ranging from about 5 to about 50 nucleotides."[128] PCT '442 Appl. (Bielas)'s methods can be used to analyze "circulating tumor mtDNA" (a known type of cfDNA) and "diagnose and stage cancer, assess response to therapy, and evaluate progression and recurrence after surgery."[129]

105. PCT '442 Appl. (Bielas) recognizes that "massively parallel sequencing represents a particularly powerful form of digital PCR" but amplification of single DNA molecules by PCR "suffers from the inherent error rate of polymerases employed for amplification, and spurious mutations generated during amplification may be misidentified as spontaneous mutations from the original (endogenous unamplified) nucleic acid."[130]

106. PCT '442 Appl. (Bielas) addresses these problems by ligating to target molecules "a single double-stranded cypher or dual cyphers (*i.e.,* barcodes or identifier tags), one on each end, so that sequencing each complementary strand can be connected back to the original molecule."[131] Thus, PCT '442 Appl. (Bielas) teaches a method of using barcodes to differentiate between mutations in the original

---

[128] *Id*. at 8:19–21.

[129] *Id.* at 25:27–26:1.

[130] *Id.* at 2:9–15.

[131] *Id.* at 7:23–27.

fragment (a "true mutation") or created by PCR (an "artifact mutation").[132]

107. PCT '442 Appl. (Bielas) explains barcodes can vary in length, "ranging from about 5 to about 50 nucleotides."[133] The barcodes on each end of a target molecule can be the same or different.[134] PCT '442 Appl. (Bielas) discloses that ligated cypher vector library is amplified by using, *e.g.*, PCR.[135] The library is then sequenced using, *e.g.*, "an Illumina® Genome Analyzer II sequencing instrument."[136]

108. After sequencing, combined Cypher-fragment sequences are used for de-duplication, as shown, e.g., in Figure 3B:

---

[132] *Id.* at 4:18–19.

[133] *Id.*, 3:4–6, 8:19–21, 14:3–16.

[134] *Id.*, 3:9–12, 4:24–27, 16:3–26.

[135] *Id.*, 24:25–28.

[136] *Id.*, 25:4–6.

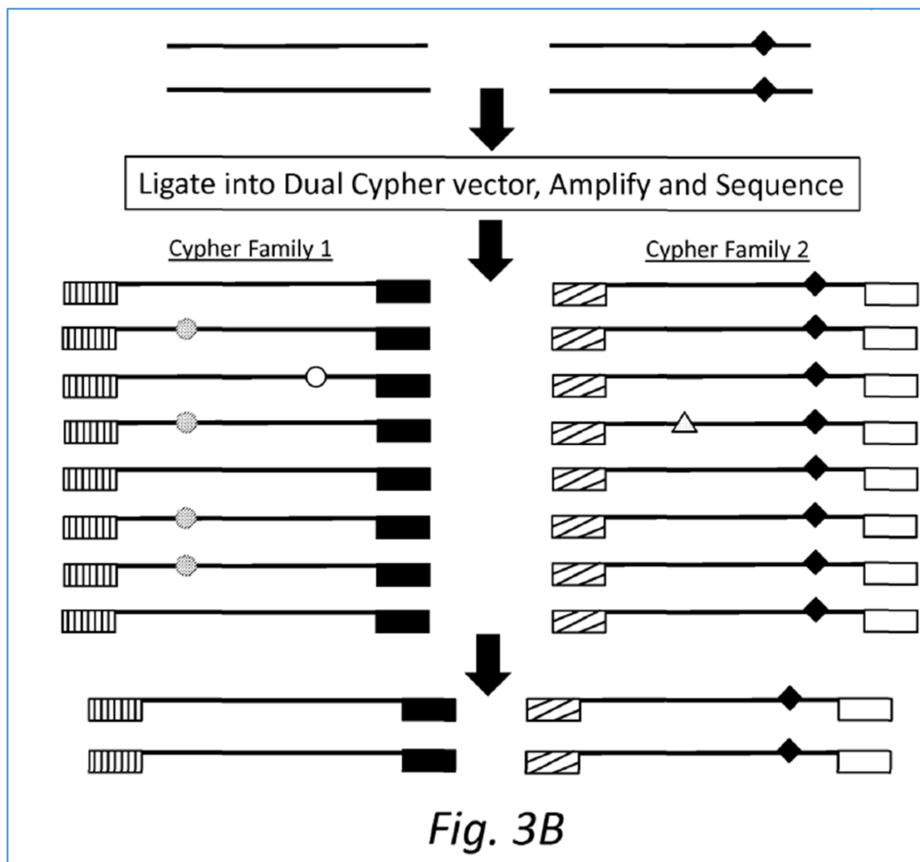


Fig. 3B

109. As taught by PCT ’442 Appl. (Bielas), “cypher tags are used to computationally deconvolute the sequencing data and map all sequence reads to single molecules (i.e., distinguish PCR and sequencing errors from real mutations).”[137] “Comparison of family sequences allowed for generation of a consensus sequence wherein ‘mutations’ (errors) arising during library preparation (open circle [in Fig. 3B]) and during sequencing (gray circle and triangle [in Fig.

---

[137] *Id.* at 25:7–9.

3B]) were computationally eliminated."[138] This process of deconvoluting sequences involves collapsing PCR copies into an original fragment sequence.[139]

B. PCT Publication No. WO 2012/142213 A2 ("PCT '213 Appl. (Vogelstein)")

110. PCT '442 Appl. (Vogelstein) is a patent application filed on April 12, 2012 and published on October 18, 2012. I understand that it is prior art under at least §102(a)(1).

111. Similar to PCT '442 Appl. (Bielas), PCT '213 Appl. (Vogelstein) discloses methods for reducing errors in parallel nucleic acid sequencing.[140] It teaches use of unique identifiers (UID), attached to one or both ends of nucleic acid fragments, that are from 2 to 4,000, from 100 to 1000, from 4 to 400, bases in length.[141] Tagged fragments are amplified to produce multiple "daughter" copies.[142]

112. PCT '213 Appl. (Vogelstein) teaches "[a]mplification of fragments containing a UID can be performed according to known techniques," including through "[p]olymerase chain reaction."[143] "Family members are sequenced and

---

[138] *Id.* at 4:12–15.

[139] *Id.* at 29:27–29.

[140] See Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶¶ [02], [05].

[141] *Id.* at ¶¶ [07], [22].

[142] *Id.* at ¶ [19].

[143] *Id.* at ¶ [27].

compared to identify any divergencies within a family. Sequencing is preferably performed on a massively parallel sequencing platform, many of which are commercially available."[144] After sequencing, "[h]igh quality reads [are] grouped into UID-families based on their endogenous or exogenous UIDs."[145] "If a mutation pre-existed in the template molecule used for amplification, that mutation should be present in a certain proportion, or even all, of daughter molecules containing that UID."[146]

C. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA ("Forshew (2012)")

113. Forshew (2012) is an article published in *Scientific Translational Medicine* on May 30, 2012. I understand that it is prior art under at least 35 U.S.C. § 102(a)(1).

114. Forshew (2012) teaches a method of detecting rare "cancer mutations present in circulating [cell-free] DNA" using a "liquid biopsy" approach.[147] Forshew (2012) teaches targeting DNA fragments from genomic regions, amplifying and sequencing the fragments, and aligning the sequences to a reference

[144] *Id.* at ¶ [28].

[145] *Id.* at ¶ [61].

[146] *Id.* at ¶ [19].

[147] *See* Ex. 1010 (Forshew (2012)) at Abstract.

sequence to identify mutations.[148] To decrease sampling errors, Forshew (2012) employs a "two-step amplification process" where the second step is selective enrichment.[149]

D. PCT Publication No. WO 2012/099832 A2 ("PCT '832 Appl. (Hendricks)")

115. PCT '832 Appl. (Hendricks) is a patent application filed on January 17, 2012, and published on July 26, 2012. I understand that it is prior art under at least 35 U.S.C. §102(a)(1).

116. PCT '832 Appl. (Hendricks) is directed to ligation reagents and methods, including for ligating nucleic acids to one another.[150] PCT '832 Appl. (Hendricks) describes "small footprint ligase" ("SFL") that ligates short polynucleotides, such as primers and probes, to other longer polynucleotides.[151]

117. PCT '832 Appl. (Hendricks) discloses how to calculate the efficiency of ligation reactions,[152] as well as factors affecting ligation efficiency.[153] Hendricks further teaches that "[t]he ligation should produce a significant or detectable amount

---

[148] *Id.* at Fig. 1, Supplement 1–4.

[149] *Id.* at p.3.

[150] *See* Ex. 1008 (PCT '832 Appl. (Hendricks)) at ¶ [0003].

[151] *Id.* at ¶ [0011].

[152] *Id.* at ¶¶ [0017], [0076].

[153] *Id.* at ¶¶ [0007], [0079].

of ligation product," and teaches achieving "10%, 20%, 30%, 50%, 60%, 70%, 75%, 80%, 85%, 90% or 95%" ligation efficiency.[154]

E. U.S. Patent Publication No. 2014/0296081 (the "'081 Patent Appl. (Diehn)")

118. The '081 Patent Appl. (Diehn) is a U.S. Patent application filed on March 13, 2014, and published on October 2, 2014. It claims priority to provisional application no. 61/798,925 (Ex. 1025, "Diehn Provisional"), which was filed March 15, 2013.

119. The Diehn Provisional specification is substantially the same as Diehn's specification, the two applications have identical claim sets, and the Diehn Provisional provides written description support for at least one claim in the '081 Patent Appl. (Diehn).[155] I understand that the '081 Patent Appl. (Diehn) is prior art under at least 35 U.S.C. §102(a)(2).

120. The '081 Patent Appl. (Diehn) is directed to methods for creating a library of mutated genomic regions and using the library to analyze cancer-specific and patient-specific genetic alterations.[156] Libraries were prepared using "cfDNA

---

[154] *Id.* at ¶ [0076].

[155] *See* Ex. 1009 ('081 Patent Appl. (Diehn)); *see also* Ex. 1025.

[156] *See* Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶¶ [0011]–[0015].

and shorn tumor, germline, and cell line genomic DNA."[157] The libraries "were constructed using the KAPA Library Preparation Kit (Kapa Biosystems)," where the DNA was washed and "eluted into 50 uL 1x ligation buffer with ligase and 100-fold molar excess of indexed Illumina TruSeq adapters."[158] Then, ligation was performed and the "ligated fragments were then amplified using 500 nM Illumina backbone oligonucleotides and a variable number of PCR cycles (between 4 and 9)."[159]

F. U.S. Patent No. 9,404,156 (the "'156 Patent (Hicks)")

121. The '156 Patent (Hicks) is a U.S. Patent that was filed on October 21, 2011 and issued on August 2, 2016. I understand that it is prior art under at least 35 U.S.C. §102(a)(2).

122. The '156 Patent (Hicks) discloses "[a] method for obtaining from genomic material genomic copy number information unaffected by amplification distortion."[160] The '156 Patent (Hicks) teaches tagging nucleic acid molecules such that the molecules become unique based on "the combination of the information in the tag and the information in the nucleic acid, which can later be read by

---

[157] *Id.* at ¶ [0142].
[158] *Id.* at ¶ [0143].
[159] *Id.* at ¶ [0143].
[160] *See* Ex. 1011 ('156 Patent (Hicks)) at Abstract.

sequencing."[161] "[S]equences are then mapped to specific regions of the human genome and deconvoluted using the unique sample tag to determine which sequence reads originated from which tumor cells."[162] "After deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted."[163]

123. For example, the '156 Patent (Hicks) discloses a counting method.

> Therefore, for a given region comprising multiple locations, the maximum copy number of the region is not less than the maximum number of different tagged nucleic acid molecules mapped to any location in that region. The maximum number of different tagged nucleic acid molecules centered about a moving window of a fixed number of locations can be taken as the measurement of true copy number of the window. This will provide at worst an underestimate of the maximum true copy number for that window, and never an overestimate.[164]

X. OPINIONS REGARDING THE VALIDITY OF THE '306 PATENT CLAIMS

A. Ground 1: Claims 1-2, 7-8, 12-13, 17-18, 20-21, 24-26, and 29 Are Rendered Obvious By PCT '442 Appl. (Bielas) in View of PCT '213 Appl. (Vogelstein)

---

[161] *Id.* at 19:11–17.

[162] *Id.* at 18:10–13.

[163] *Id.* at 19:36–38.

[164] *Id*. at 19:44–53.

130. For the reasons set forth *infra*, Challenged Claims 1-2, 7-8, 12-13, 17-18, 20-21, 24-26, and 29 of the ’306 Patent are invalid as obvious over PCT ’442 Appl. (Bielas) in view of PCT ’213 Appl. (Vogelstein).

1. Motivation To Combine

131. In my opinion, one of ordinary skill in the art would have been motivated to combine the teachings of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein).

132. PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) are analogous art directed to the field of nucleic acid sequencing that those of ordinary skill in the art would have looked to when screening cfDNA for mutations.[165] Both references recognize that it was desirable to identify mutations that are present in small fractions of DNA using massively parallel sequencing (*e.g.*, for detecting molecular residual disease (MRD) following surgical resections, as well as detecting rare genetic variants associated with cancer and those mutations that confer resistance to),[166] but that high error rates made this challenging.[167] Both references

---

[165] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 1:15–19; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [02].

[166] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 24:4–11; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [03].

[167] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 2:7–22; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶¶ [04]–[05].

disclose methods to address the errors, and thus improve detection accuracy, by attaching barcodes to the target molecules prior to amplification. For example, PCT '213 Appl. (Vogelstein) discloses methods of sequencing where exogenous barcodes are attached to one or both ends of molecules of interest and then used those barcodes to group sequence reads and distinguish "true mutations" from errors.[168]

133. In PCT '213 Appl. (Vogelstein), fragments are tagged with a "Unique Identifier (UID)" and amplified to form families of fragments that each share the same UID.[169] The family members are then "sequenced and compared to identify any divergencies within a family."[170] Thus, PCT '213 Appl. (Vogelstein) discloses grouping sequence reads into single-stranded (unpaired) families. PCT '213 Appl. (Vogelstein) recognized, however, that "*specificity can be even further increased by requiring that each strand of the original double stranded template contain the mutation* . . . ."[171] Thus, PCT '213 Appl. (Vogelstein) teaches that considering families of paired reads could further reduce errors.

134. PCT '442 Appl. (Bielas) discloses attaching barcodes to each end of duplex strands of target DNA molecules, amplifying and sequencing the molecules,

---

[168] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶¶ [19], [23].

[169] *Id*. at ¶¶ [19], [27].

[170] *Id*. at ¶ [28].

[171] *Id*. at ¶ [67] (emphasis added).

and grouping sequencing reads along with their reverse complements into paired families to distinguish mutations from errors.[172] According to PCT '442 Appl. (Bielas), "mutations that are present in all or substantially all reads [] from the same cypher and its reverse complement are counted as true mutations."[173] PCT '442 Appl. (Bielas) also explained that "[c]omparison of family sequences allowed for generation of a consensus sequence wherein 'mutations' (errors) arising during library preparation [] and during sequencing [] were computationally eliminated."[174]

135. One of ordinary skill in the art would have been motivated to combine the teachings of PCT '442 Appl. (Bielas) with the teachings of PCT '213 Appl. (Vogelstein) for several reasons. Both references are in the same field of DNA sequencing to identify rare genetic variants in cancer, recognize the benefits of massively parallel sequencing to identify such variants, and the associated problems from introduced errors. Both references also sought to reduce errors associated with the sequencing workflow by tagging target fragments with barcodes that are used to group families of sequence reads and generate consensus sequences. Moreover, given PCT '213 Appl. (Vogelstein)'s recognition of the benefits of comparing

---

[172] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:9–12.

[173] *Id*. at 5:15–17.

[174] *Id*. at 4:12–15.

complementary strands, as described *supra*, and PCT ’442 Appl. (Bielas)’ teaching of methods for reducing errors by using barcodes to group sequencing reads with their complements, one of ordinary skill in the art would have been motivated to consider the teachings of Bielas in connection with the teachings of Vogelstein in order to further reduce or eliminate errors in sequencing, and thus improve variant detection.

136. One of ordinary skill in the art combining PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) would have had a reasonable expectation of success in achieving the claimed inventions. The references both teach that their methods are general techniques for NGS and error correction, which can be used with standard and well-known methods for library preparation, tagging, amplification and sequencing. Both references also rely on standard molecular biology tools, and neither requires novel chemistry or unconventional instrumentation. As discussed *supra*, the sequencing methods in PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) were well-known in the art by December 2013. One of ordinary skill in the art would have understood that the methods disclosed in PCT ’213 Appl. (Vogelstein) could be used with those disclosed in PCT ’442 Appl. (Bielas).

137. In particular, one of ordinary skill in the art would have understood that PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein)’s grouping into families, reducing redundancy, and mapping to a reference sequence could be successfully

applied to both paired reads and unpaired reads, as both references teach the advantages of generating families of sequencing reads using barcodes to reduce errors. Further, one of ordinary skill in the art would have expected that PCT '442 Appl. (Bielas)' methods (applied to circulating tumor DNA) could be successfully used in conjunction with PCT '213 Appl. (Vogelstein)'s methods (applied to DNA in blood and maternal plasma), as these types of DNA are all present in small amounts in natural samples.

2. Independent Claim 1

(a) Claim 1[pre]: A method, comprising:

138. To the extent the preamble is considered limiting, it is my opinion that PCT '442 Appl. (Bielas) discloses this limitation.

139. PCT '442 Appl. (Bielas) discloses methods for detecting mutations in DNA sequences.[175]

(b) Element [1a]: "(a) providing a population of cell-free deoxyribonucleic acid (cfDNA) molecules having first and second complementary strands;"

140. It is my opinion that PCT '442 Appl. (Bielas) discloses this limitation.

141. For example, PCT '442 Appl. (Bielas) discloses that its methods are "useful in detecting rare mutants against a large background signal, such as when . .

---

[175] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 7:23–27.

. detecting *circulating mutant DNA in blood*,"[176] and "to quantify circulating tumor cells (CTCs) and *circulating tumor mtDNA (ctmtDNA)* . . . ."[177] PCT '442 Appl. (Bielas) also discloses DNA of various lengths and sources, including "*double-stranded nucleic acid molecules that may be fragments* or shorter molecules generated from longer nucleic acid molecules, including from natural samples."[178] Circulating tumor DNA (ctDNA) were a known type of cfDNA. They are fragments of DNA freely floating in the bloodstream, outside of cancer cells, contain cancer mutations and exist within a large population of normal cfDNA.

(c) Element [1b1]: "(b) tagging a plurality of the cfDNA molecules in the population with duplex tags comprising molecular barcodes to produce tagged parent polynucleotides, wherein the duplex tags are attached to both ends of a molecule of the plurality of the cfDNA molecules,"

142. It is my opinion that PCT '442 Appl. (Bielas) discloses this limitation.[179]

---

[176] *Id*. at 23:12–14 (emphasis added).

[177] *Id*. at 25:27–28 (emphasis added).

[178] *Id*. at 8:30–9:1 (emphasis added).

[179] *See* Ex. 1001 ('306 Patent) at 17:9–13, which discloses that "duplex tags" are "tags that differently label the complementary strands (i.e., the 'Watson' and 'Crick' strands) of a double-stranded molecule."

143. PCT ’442 Appl. (Bielas) discloses “a double-stranded nucleic acid library wherein target nucleic acid molecules include dual cyphers *(i.e.,* barcodes or origin identifier tags), one on each end (same or different), so that sequencing each complementary strand can be connected or linked back to the original molecule.”[180] The dual cyphers (i.e., molecular barcodes) disclosed in PCT ’442 Appl. (Bielas) are “duplex tags” because they differentially label the complimentary strands of a double-stranded cfDNA molecule. The barcodes in PCT ’442 Appl. (Bielas) are double-stranded nucleic acid molecules comprised of basepairs. For example, PCT ’442 Appl. (Bielas) explains that “[t]arget nucleic acid molecules were ligated into a cypher vector library containing previously catalogued *dual, double-stranded cyphers*.”[181] PCT ’442 Appl. (Bielas) also explains that “[a] single [NGS] run on MiSeq® demonstrated optimal coverage and diversity at the upstream seven *basepair* cypher in the vector library.”[182] One of ordinary skill in the art would have understood that a “base-pair” consists of two different, but complementary, nucleotide bases (*e.g.*, adenine and thymine) that are paired together. Because the nucleotide bases are complementary, one of ordinary skill in the art would have

---

[180] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 4:24–27; *see also id.* at 8:19–28.

[181] *Id*. at 4:9–10 (emphasis added).

[182] *Id*. at 4:9–10 (emphasis added); *see also id.* at 7:25–26, 12:11–21, 27:27–28.

understood that the nucleotide bases in the top strand are different than the nucleotide bases on the bottom strand. Thus, the dual cyphers in PCT '442 Appl. (Bielas) are duplex tags comprising molecular barcodes.

144. The barcodes in PCT '442 Appl. (Bielas) are attached at each end of the target molecules.[183] PCT '442 Appl. (Bielas) provides examples of barcodes attached to both ends of the DNA molecule such that the molecules have, *e.g.*, "a formula of ... X^a_Y_X^b ... or X^b_Y_X^a (in 5' to 3' order), wherein (a) X^a comprises a first random cypher, (b) Y comprises a target nucleic acid molecule, and (c) X^b comprises a second random cypher."[184] One of ordinary skill in the art would have understood that the tagged polynucleotides in PCT '442 Appl. (Bielas) are tagged ***parent*** polynucleotides because they are subsequently amplified and sequenced to generate progeny polynucleotides.[185]

(d) Element [1b2]: "wherein the plurality of the cfDNA molecules are tagged with n different combinations of molecular barcodes, wherein n is at least 2 and no more than 100,000*z, wherein z is a mean of an expected number of duplicate molecules in the population of cfDNA molecules that map to identical start and stop positions on a reference sequence;"

[183] *Id*. at 16:12–13.

[184] *Id*. at 2:29–3:4; *see also id*. at 12:11–21; Figs. 2, 3A, 3B.

[185] *Id.* at 16:11–13; *see also id.* at 4:24–26, 8:26–28, 12:11–20, 15:12–15.

145. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this limitation.

146. Without taking a position on whether one of ordinary skill in the art would have understood the “z” term, based on Claim 7 of the ’306 Patent, the “z” term must include 2 to 8. Under this interpretation, PCT ’442 Appl. (Bielas) alone or in view of PCT ’213 Appl. (Vogelstein) discloses this limitation.

147. The tags in PCT ’442 Appl. (Bielas) have “a length ranging from about 5 to about 50 nucleotides.”[186] The barcodes at each end in PCT ’442 Appl. (Bielas) together form a combination of molecular barcodes. PCT ’442 Appl. (Bielas) explains that “[t]he number of nucleotides contained in each of the random cyphers or barcodes will govern the total number of possible barcodes available for use in a

---

[186] *Id.* at 8:19–21. I note that one of ordinary skill in the art also would have understood that a length of “about 5” nucleotides includes a length of 4 nucleotides. At a minimum, it would have been obvious in view of PCT ’213 Appl. (Vogelstein) to use barcodes of a length of 4 nucleotides. Such barcodes would yield a maximum of 256 (*i.e.*, 4^4) different combinations if the same barcode were used on each end of the target molecule, or 65,536 different combinations (*i.e.*, 4^4 x 4^4 (or 256 x 256)) if a different barcode were used on each end. As taught by PCT ’442 Appl. (Bielas) at 14:10–13, however, it may be desirable to use barcodes that “exclud[e], for example, sequences in which all the nucleotides are identical *(e.g.,* all A or all T or all C or all G) or [] exclud[e] sequences in which three contiguous nucleotides are identical or [] exclud[e] both of these types of molecules.” This exclusion of specific sequence types would result in a smaller number of combinations within the range of 2 to 100,000*z.

library."[187] PCT '442 Appl. (Bielas) also explains "[f]or example, a barcode of 7 nucleotides would have a formula of 5'-NNNNNNN-3' . . . wherein N may be any naturally occurring nucleotide. The four naturally occurring nucleotides are A, T, C, and G, so the total number of possible random cyphers is 4^7, or 16,384 possible random arrangements (*i.e.*, 16,384 different or unique cyphers)."[188]

148. PCT '442 Appl. (Bielas) further discloses the number of different combinations of barcodes depends upon whether the barcode at the 5' end is the same or different than the barcode at the 3' end. For example, PCT '442 Appl. (Bielas) discloses that molecules may be tagged with the same barcode on both ends. "In certain embodiments, the double-stranded sequences of the X^a and X^b cyphers are the same (*e.g.*, $X^a = X^b$) for one or more target nucleic acid molecules, ..."[189] In such cases, the number of combinations of barcodes is simply four (4) to the power of the barcode length (*i.e.*, $4^{[\text{barcode length}]}$).[190]

---

[187] *Id*. at 13:17–18; *see also id.* at 13:18–20.

[188] *Id*. at 14:3–7.

[189] *Id*. at 3:9–10; *see also id.* at 5:28–30.

[190] *Id*. at 14:3–8.

149. For example, if the barcodes are 5 nucleotides in length, there are 4^5 or 1,024 possible barcode combinations.[191] Thus, one of ordinary skill in the art would have understood that PCT '442 Appl. (Bielas)'s tags of "about 5" nucleotides are within the claimed range of 2 to no more than 100,000*z regardless of the value of z. In the particular case that the value of z is equal to a number between 2 and 8 (of the number of barcodes "n" would be from 2 to between 200,000 and 800,000).

150. To the extent further disclosure is required, PCT '213 Appl. (Vogelstein) discloses that molecular barcodes—its "unique identifiers" or "UIDs"[192]—"may be added to one or both ends of the fragments."[193] PCT '213 Appl. (Vogelstein) also discloses the barcodes "consist[] of 2 to 4,000 nucleotides"[194] and may consist of "at least 4, at least 16, at least 64, at least 256 . . . different sequences."[195] Using PCT '213 Appl. (Vogelstein)'s UIDs on both ends of a cfDNA molecule, as disclosed in both PCT '213 Appl. (Vogelstein) and PCT '442 Appl. (Bielas), both references disclose the claimed "plurality of the cfDNA

---

[191] I note that PCT '442 Appl. (Bielas) at 16:3–5 also discloses that "random cypher sequences from a particular pool of cyphers (*e.g.*, pools of 4,094, 16,384 or 65,536 unique cyphers) may be used more than once."

[192] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [07].

[193] *Id*. at ¶ [22].

[194] *Id*.

[195] *Id*. at ¶ [30].

molecules are tagged with n different combinations of molecular barcodes, wherein n is at least 2 and no more than 100,000*z . . ." For example, barcodes of 2 nucleotides would have 4^2 x 4^2 = 256 possible barcode combinations.

(e) Element [1c]: "(c) amplifying a plurality of the tagged parent polynucleotides to produce amplified progeny polynucleotides;"

151. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this limitation.

152. PCT '442 Appl. (Bielas) discloses that after "[t]arget nucleic acid molecules were ligated into a cypher vector library containing previously catalogued dual, double-stranded cyphers. The target sequences were amplified and sequenced."[196] Techniques for amplifying DNA molecules were well-known in the art, as described *supra*, and PCT '442 Appl. (Bielas) discloses that the molecules can be amplified using such known techniques.[197] This process produces amplified progeny polynucleotides. PCT '442 Appl. (Bielas) explains that "[NGS] has opened the door to sequencing multiple copies of an amplified single nucleic acid molecule - referred to as deep sequencing."[198]

---

[196] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:9–11; *see also id.* at 3:17–22.

[197] *See*, *e.g.*, *id*. at 5:30–6:3.

[198] *Id*. at 7:8–9; *see also id.* at 6:11–13.

153. To the extent further disclosure is required, PCT ’213 Appl. (Vogelstein) also discloses amplifying a plurality of the tagged parent polynucleotides to produce progeny polynucleotides. For example, PCT ’213 Appl. (Vogelstein) discloses that “[a]mplification of fragments containing a UID can he performed according to known techniques to generate families of fragments,” including through PCR.[199] PCT ’213 Appl. (Vogelstein) also discloses that “[a]mplification forms a family of fragments, each member of the family sharing the same UID.”[200] One of ordinary skill in the art would have understood that the family of fragments are progeny polynucleotides.

(f) Element [1d]: “(d) sequencing at least a subset of the amplified progeny polynucleotides to produce a set of sequence reads; and”

154. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this limitation.

155. PCT ’442 Appl. (Bielas) discloses that “[e]ach strand of the double-stranded nucleic acid library (*e.g.*, genomic DNA, cDNA) can be amplified and sequenced using, for example, [NGS] technologies . . . .”[201] NGS techniques were

---

[199] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [27].

[200] *Id*.

[201] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 5:30–6:3; *see also id.* at 10:25–27.

well-known in the art.[202] Moreover, PCT '442 Appl. (Bielas) also discloses that the molecules can be sequenced using such known techniques.[203] The sequencing process generates "sequence information from each complementary strand of a first double-stranded nucleic acid molecule . . . "[204] One of ordinary skill in the art would have understood this passage to mean *sequence reads*.

156. To the extent further disclosure is required, PCT '213 Appl. (Vogelstein) also discloses this element. For example, PCT '213 Appl. (Vogelstein) explains that "[s]equencing is preferably performed on a massively parallel sequencing platform, many of which are commercially available."[205]

(g) Element [1e]: "(e) reducing or tracking redundancy of a plurality of sequence reads from the set of sequence reads using at least sequencing information from the molecular barcodes of the duplex tags to determine distinct cfDNA molecules from among the tagged parent polynucleotides, wherein the distinct cfDNA molecules are determined based on (i) paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides, or (ii) unpaired reads corresponding to sequence reads generated from a first tagged strand having no second tagged complementary

---

[202] *See supra* at § VII.E.

[203] *Id.* at 10:27–29.

[204] *Id.* at 6:3–4.

[205] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [28].

strand derived from cfDNA molecules from among the tagged parent polynucleotides,"

157. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this limitation.[206]

158. PCT '442 Appl. (Bielas) discloses that "[a]ll sequencing reads having identical cypher pairs, along with their reverse complements, were grouped into families. One of ordinary skill would have understood this to mean that redundant reads (*i.e.*, amplified copies of the parent polynucleotide) are grouped into families based at least on sequence information from the barcodes because they are grouped based on having "identical cypher pairs." Because the sequence reads in PCT '442 Appl. (Bielas) are grouped in families "along with their reverse complements," they are families of paired reads. Once the paired sequence reads are grouped into families, PCT '442 Appl. (Bielas) discloses that redundancy is reduced and/or tracked to determine distinct cfDNA molecules by generating consensus sequences. For example, PCT '442 Appl. (Bielas) discloses that "[c]omparison of family sequences allowed for generation of a consensus sequence wherein 'mutations' (errors) arising during library preparation (open circle) and during sequencing (gray

---

[206] *See* Ex. 1028 at pp. 24–25. I note in the IPR2022-01400 decision, the Board determined that the plain language of this claim element requires use of paired ***or*** unpaired reads, but not both.

circle and triangle) were computationally eliminated."[207] This is shown *infra* in Figure 3B of PCT '442 Appl. (Bielas).[208]

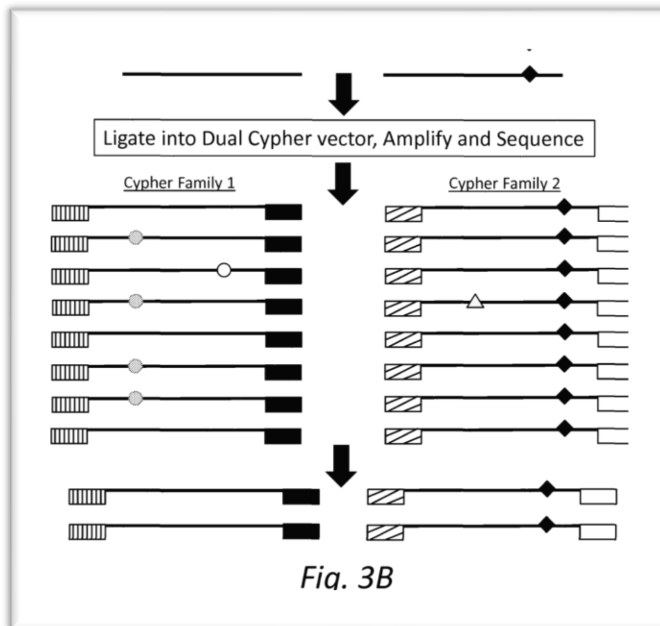


Fig. 3B

As shown in the figure *supra*, PCT '442 Appl. (Bielas) teaches using families of paired reads to reduce and/or track redundancy of a plurality of sequence reads by generating a consensus sequence that identifies a distinct cfDNA molecule.[209]

159. To the extent additional disclosure is required, PCT '213 Appl. (Vogelstein) discloses reducing and/or tracking redundancy based on families of unpaired reads. For example, PCT '213 Appl. (Vogelstein) discloses "amplification

---

[207] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:11–15.

[208] *Id* at pp. 40–41.

[209] *Id* at 4:11–15; *see also id*. at 23:6–11.

of each uniquely tagged template, so that many daughter molecules with the identical sequence are generated . . . ."[210] After sequencing, "[h]igh quality reads [are] grouped into UID-families based on their endogenous or exogenous UIDs."[211] PCT '213 Appl. (Vogelstein) also discloses that "[a] nucleotide sequence is identified as accurately representing an analyte nucleic acid fragment when at least 1% of members of the family contain the sequence."[212] Thus, one of ordinary skill in the art would have understood that redundant reads are reduced to identify an "accurate[] . . . analyte nucleic acid fragment" by forming families of unpaired reads using at least sequencing information from the barcodes.

160. Furthermore, PCT '213 Appl. (Vogelstein) recognizes that "various amplifications are not perfect, so every strand of every original template molecule is not recovered as a UID family."[213] Accordingly, one of ordinary skill in the art would have understood that at least some reads in a family have no second tagged complementary strand derived from among the tagged parent polynucleotides. As described *supra*, only a fraction of the library fragments from the original library are used in the bridge amplification step, resulting in some sequence read having no

---

[210] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [19].

[211] *Id*. at ¶ [61].

[212] *Id*. at ¶ [07].

[213] *Id*. at ¶ [66].

second tagged complementary strand. Accordingly, one of ordinary skill in the art would have understood that at least some of the families in PCT ’213 Appl. (Vogelstein) are families of unpaired reads that have no second tagged complementary strand derived from among the tagged parent polynucleotides.

161. The fact that not all sequence reads would have a complementary strand would have motivated one of ordinary skill in the art to combine the teachings of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein). Although PCT ’213 Appl. (Vogelstein) primarily discloses grouping sequence reads into unpaired families, it recognizes that “specificity can be even further increased by requiring that each strand of the original double stranded template contain the mutation.”[214] In other words, PCT ’213 Appl. (Vogelstein) suggests that considering families of paired reads could further reduce errors when reducing redundancy, and PCT ’442 Appl. (Bielas) teaches a method for using families of paired reads to reduce redundancy. Although Claim 1 of the ’306 Patent only requires families of either paired *or* unpaired reads, the combination of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) teaches reducing redundancy based on both paired and unpaired reads.

(h) Element [1f]: “wherein reducing or tracking the redundancy of the plurality of sequence reads

[214] *Id*. at ¶ [67]; *see also id*. at ¶ [42].

comprises mapping at least a subset of the plurality of sequence reads to the reference sequence."

162. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this limitation.

163. For example, PCT '442 Appl. (Bielas) discloses that "using the methods of this disclosure, a person of ordinary skill in the art can associate each DNA sequence read to an original template DNA."[215] PCT '442 Appl. (Bielas) discloses that "[b]ase calling and sequence alignment will be performed using, for example, the Eland pipeline (Illumina, San Diego, CA)."[216] One of ordinary skill in the art would have known that the ELAND pipeline is computational tool used for processing raw sequencing data generated by Illumina's NGS platforms and mapping the sequence reads to a reference sequence.[217] Thus, the base calling and sequence alignment in PCT '442 Appl. (Bielas) involves mapping sequencing reads to a reference sequence in connection with reducing redundancy by generating a consensus sequence.

164. To the extent additional disclosure is required, PCT '213 Appl. (Vogelstein) also discloses that after sequencing "[b]ase-calling and sequence

---

[215] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 23:6–7.

[216] *Id*. at 25:9–10.

[217] *See* Ex. 1061 (CASAVA 1.8 User Guide) at 6; *see also* § VII.I *supra*.

alignment was performed with the Eland pipeline (Illumina)."[218] PCT '213 Appl. (Vogelstein) also discloses that after mapping the sequence reads "[o]nly high quality reads . . . were used for subsequent analysis."[219] PCT '213 Appl. (Vogelstein) then explains that these "[h]igh quality reads were grouped into UID-families based on their endogenous or exogenous UIDs."[220] Thus, one of ordinary skill in the art would have understood that PCT '213 Appl. (Vogelstein) discloses mapping at least a subset of the plurality of sequence reads to the reference sequence as part of the process of reducing or tracking the redundancy of the plurality of sequence reads.

3. Claim 2: "The method of claim 1, wherein the population of cfDNA molecules is obtained or derived from a subject having cancer."

165. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

166. PCT '442 Appl. (Bielas) discloses "target nucleic acid molecules . . . refer to a plurality of double-stranded nucleic acid molecules that may be fragments or shorter molecules generated from longer nucleic acid molecules, including from

---

[218] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [60].

[219] *Id*.

[220] *Id*. at ¶ [60].

natural samples (e.g., a genome) . . ."[221] PCT '442 Appl. (Bielas) further discloses the methods disclosed "will be useful in detecting rare mutants against a large background signal, such as when . . . detecting circulating mutant DNA in blood, monitoring or detecting disease and rare mutations by direct sequencing . . ."[222] PCT '442 Appl. (Bielas) also discloses the use of "circulating tumor mtDNA (ctmtDNA)," which can be quantified to "*diagnose and stage cancer*."[223] PCT '442 Appl. (Bielas) also discloses that after "mutations from individual tumors are identified, patient-matched blood specimens will be examined for the presence of identical mutations in the plasma and buffy coat to determine the frequencies of ctmtDNA and CTCs . . ."[224] A person of ordinary skill in the art would understand that "circulating mutant DNA in blood" and "ctmtDNA" are cfDNA. Thus, PCT '442 Appl. (Bielas) discloses "cfDNA molecules [] obtained or derived from a subject having cancer."

---

[221] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 8:29–9:1.

[222] *Id*. at 23:12–16.

[223] *Id*. at 25:15–26:12 (Example 2, emphasis added); *see also id.* at 24:4–20.

[224] *Id*. at 26:6–9.

4. Claim 7: "The method of claim 1, wherein z is between 2 and 8."

167. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein discloses this Claim element.

168. Where z is between 2 and 8, there are at least 2 and no more than 200,000 to 800,000 different combinations of molecular barcodes. As discussed in connection with Element [1c] of the '306 Patent,[225] PCT '442 Appl. (Bielas) discloses molecular barcodes "having a length ranging from about 5 to about 50 nucleotides" and also that the molecules may be tagged with the same barcode on both ends.[226] As described *supra*, when the barcodes are 5 nucleotides in length, and the same barcode is used on each end, there are 4^5 or 1,024 possible barcode combinations, which falls within the claimed range. PCT '442 Appl. (Bielas) also discloses further reducing the number of barcode combinations by excluding, *e.g.*, barcodes that contain three contiguous nucleotides that are identical.[227]

169. To the extent further disclosure is needed, PCT '213 Appl. (Vogelstein) discloses wherein z is between 2 and 8. As discussed in connection with Element

---

[225] *See supra* at § X.A.2.

[226] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 8:19–21; *see also id*. at 3:9–10.

[227] *Id.* at 14:17-28.

[1c] of the ’306 Patent, *supra*,[228] PCT ’213 Appl. (Vogelstein) discloses molecular barcodes “consisting of 2 to 4,000 nucleotides”[229] and a range of numbers of different combinations of molecular barcodes, including “at least 4, at least 16, at least 64, at least 256.”[230] Thus, one of ordinary skill in the art would have understood that barcodes of 2-5 nucleotides, or combinations of barcodes of “at least 4” to “at least 256” discloses the claimed range.

5. Claim 8: “The method of claim 1, wherein the molecular barcodes have a length of 5 to 20 base pairs.”

170. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.[231]

171. PCT ’442 Appl. (Bielas) discloses molecular barcodes with “a length ranging from about 5 nucleotides to about 50 nucleotides (or about 5 nucleotides to . . . about 20 nucleotides).”[232] The barcodes in PCT ’442 Appl. (Bielas) are double-stranded, which one of ordinary skill in the art would have understood means they

---

[228] *See supra* at § X.A.2.

[229] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [22].

[230] *Id*. at ¶ [30].

[231] I note that one of ordinary skill in the art would have understood that the claimed “base-pair” consists of two complementary nucleotide bases that are paired together. *See also supra* at § VII.A.

[232] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 3:4–8.

consist of base-pairs of nucleotides.[233] Thus, PCT ’442 Appl. (Bielas) discloses the claimed range of “5 to 20 base pairs.”

172. To the extent further disclosure is needed, PCT ’213 Appl. (Vogelstein) discloses molecular barcodes “consisting of 2 to 4,000 nucleotides.”[234] One of ordinary skill in the art would have understood that PCT ’213 Appl. (Vogelstein)’s range of “2 to 4,000 nucleotides” entirely encompasses the claimed range of “5 to 20 base pairs.”

6. Claim 12: “The method of claim 1, wherein the duplex tags are part of sequencing adapters.”

173. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.

174. PCT ’442 Appl. (Bielas) discloses an exemplary figure of a “vector of the present disclosure, *wherein adaptor sequences are included* and are useful for, for example, bridge amplification methods before sequencing.”[235]

---

[233] *Id*. at 4:9–10; *see also id*. at 27:27–28.

[234] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [22].

[235] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 4:1–3 (emphasis added).

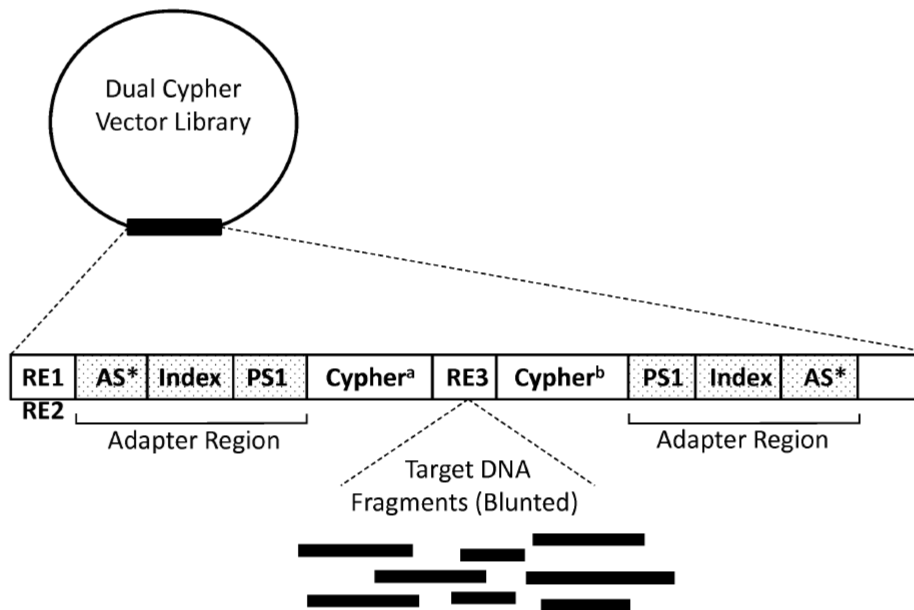


Fig. 2

175. PCT '442 Appl. (Bielas) discloses that the "DNA fragments are ligated into the *Sma*I site of the library of dual cypher vectors shown in Figure 2 to generate a target genomic library."[236] One of ordinary skill in the art would have understood that this discloses that the duplex tags, which PCT '442 Appl. (Bielas) calls "cyphers," are contained within sequencing adapters.[237]

176. To the extent further disclosure is needed, PCT '213 Appl. (Vogelstein) discloses that "DNA (sheared or unsheared) is amplified with a set of gene-specific primers. One of the primers has a random DNA sequence *(e.g.,* a set of 14 N's) that

---

[236] *Id*. at 24:23–24.

[237] *Id*. at Claim 14 ("wherein the plurality of random cyphers further comprise at least one adapter sequence (AS)").

forms the unique identifier (UID) (variously shaded bars), located 5' to its gene-specific sequence, and both have sequences that permit universal amplification in the next step (earth hatched and cross hatched bars)."[238] PCT '213 Appl. (Vogelstein) also discloses that "[a]ttachment of an exogeneous UID to an analyte nucleic acids fragment may be performed by any means known in the art, including enzymatic, chemical, or biologic. One means employs a polymerase chain reaction. Another means employs a ligase enzyme."[239] PCT '213 Appl. (Vogelstein) also explains that "[a] UID may be contained within a nucleic acid molecule that contains other regions for other intended functionality."[240] Thus, one of ordinary skill in the art reading a particular disclosure in PCT '213 Appl. (Vogelstein) involving PCR to attach UIDs to the target nucleic acid could have easily substituted the attachment of the UIDs by a ligation reaction such that the UIDs are part of sequencing adapters.

7. Claim 13: "The method of claim 1, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises grouping the paired reads or the unpaired reads into families based at least in part on (i) the molecular barcodes associated with the paired reads or the unpaired

---

[238] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [14].

[239] *Id*. at ¶ [22].

[240] *Id*.

reads and (ii) sequence information at the start and stop positions of the paired reads or the unpaired reads.”

177. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.

178. As discussed *supra* regarding Element [1e] of the ’306 Patent, PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) disclose reducing and/or tracking the redundancy of sequence reads by grouping paired or unpaired reads into “families” based on the molecular barcodes associated with the paired or unpaired reads.[241]

179. PCT ’442 Appl. (Bielas) further discloses reducing and/or tracking the redundancy of sequence reads by grouping paired reads into families based on sequence information at the start and stop positions of the paired reads. PCT ’442 Appl. (Bielas) discloses that “the first about 5 nucleotides to about 20 nucleotides of the target nucleic acid molecule sequence may be used as a further identifier tag together with the sequence of an associated random cypher.”[242] PCT ’442 Appl. (Bielas) further discloses using the first about 5 nucleotides to about 20 nucleotides at both ends, i.e., at the start and stop positions.[243] Using sequence information at

---

[241] *See supra* at § X.A.2(g).

[242] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 14:13–16.

[243] *Id.* at 15:5–11.

the start and stop positions, PCT ’442 Appl. (Bielas) discloses reducing and/or tracking the redundancy of sequence reads.[244]

180. To the extent further disclosure is needed, PCT ’213 Appl. (Vogelstein) also discloses reducing and/or tracking the redundancy of sequence reads by grouping paired reads or unpaired reads into “families” based on either exogenous barcodes, endogenous barcodes formed from the sequence information at the ends (*i.e.*, start and stop positions) of the target fragments, or a combination of the two.[245] PCT ’213 Appl. (Vogelstein) also explains that “[t]he resulting DNA fragments contained UIDs composed of three sequences: two endogenous ones, represented by the two ends of the original sheared fragments plus the exogenous sequence introduced during the indexing amplification.”[246] Thus, one of ordinary skill in the art would have understood from PCT ’213 Appl. (Vogelstein) that sequence reads can be grouped into families based on the barcodes, as well as based on sequence information at the start and stop positions of the sequence reads.

8. Claim 17

(a) Claim [17pre]: “A method, comprising:”

---

[244] *Id*. at 4:12–15.

[245] *See, e.g.*, Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶¶ [07], [09], [26].

[246] *Id*. at ¶ [56].

181. To the extent the preamble is considered limiting, it is my opinion that PCT ’442 Appl. (Bielas) discloses this limitation.[247]

(b) Element [17a]: “(a) tagging a population of double-stranded cell-free deoxyribonucleic acid (cfDNA) molecules obtained or derived from a sample of a subject with a set of tags comprising molecular barcodes to produce tagged parent polynucleotides;”

182. It is my opinion that PCT ’442 Appl. (Bielas) discloses this limitation.

183. Element [17a] is substantially the same as Element [1b1] of the ’306 Patent. As discussed *supra*, PCT ’442 Appl. (Bielas) discloses all the elements of Element [1b1] of the ’306 Patent.[248]

184. Element [17a] specifically recites tagging “double-stranded” cfDNA molecules “from a sample of a subject with a set of tags comprising molecular barcodes.” As discussed *supra* regarding Element [1a] of the ’306 Patent, PCT ’442 Appl. (Bielas) discloses the use of cell-free DNA (cfDNA) molecules having first and second complementary strands.[249] Also described *supra* regarding Claim 2 of the ’306 Patent, PCT ’442 Appl. (Bielas) discloses that the cfDNA molecules are obtained or derived “from a sample of a subject.”[250]

---

[247] *See* discussion *supra* of Element 1[pre] of the ’306 Patent at § X.A.2(a).

[248] *See supra* at § X.A.2(c).

[249] *See supra* at § X.A.2(b).

[250] *See supra* at § X.A.3.

(c) Element [17b]: "(b) amplifying a plurality of the tagged parent polynucleotides to produce amplified progeny polynucleotides;"

185. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element. PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses all the elements of Element [17b] for the same reasons discussed *supra* for Element [1c] of the '306 Patent.[251]

(d) Element [17c]: "(c) sequencing at least a subset of the amplified progeny polynucleotides to produce a set of sequence reads; and"

186. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element. PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses all the elements of Element [17c] for the same reasons discussed *supra* for Element [1d] of the '306 Patent.[252]

(e) Element [17d]: "(d) sorting a plurality of sequence reads from the set of sequence reads into (i) families comprising paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand derived from double-stranded cfDNA molecules from among the tagged parent polynucleotides, and (ii) families

---

[251] *See supra* at § X.A.2(d).

[252] *See supra* at § X.A.2(e).

comprising unpaired reads corresponding to sequence reads generated from a first tagged strand having no second tagged complementary strand derived from double-stranded cfDNA molecules from among the tagged parent polynucleotides."

187. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

188. PCT '442 Appl. (Bielas) discloses sorting sequence reads from the set of sequence reads into families comprising paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand. Specifically, PCT '442 Appl. (Bielas) discloses that "[a]ll sequencing reads having identical cypher pairs, along with their reverse complements, were grouped into families."[253] This is shown *infra* in Figure 3B of the PCT '442 Appl. (Bielas).[254]

---

[253] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:11–15.

[254] *Id*. at pp. 40–41.

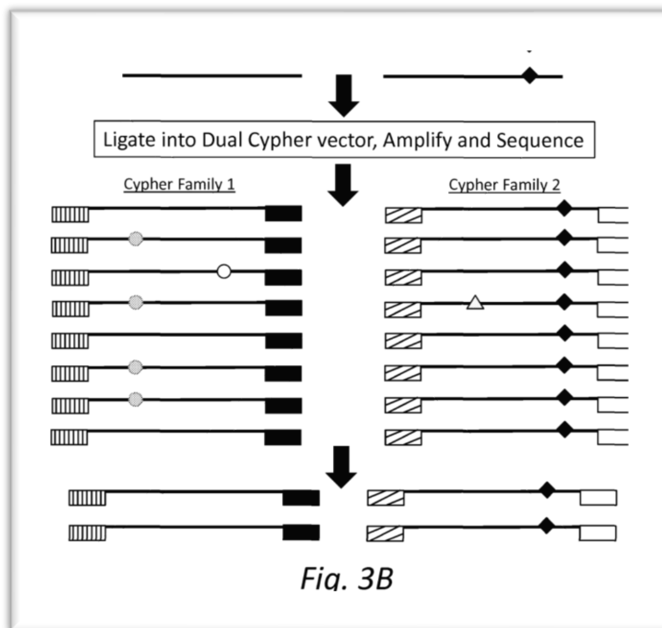


Fig. 3B

One of ordinary skill in the art would have understood that these families comprise "paired reads corresponding to sequence reads generated from a first tagged strand and a second tagged complementary strand" because PCT '442 Appl. (Bielas) uses "sequencing reads having identical cypher pairs, along with their reverse complements,"[255] indicating that these reads are paired reads.

189. PCT '213 Appl. (Vogelstein) discloses sorting sequence reads into families comprising unpaired reads. For example, PCT '213 Appl. (Vogelstein) discloses that "amplification of each uniquely tagged template, so that many daughter molecules with the identical sequence are generated (defined as a UID

[255] *Id*. at 4:11–12.

family)."[256] After sequencing, PCT '213 Appl. (Vogelstein) explains that "[h]igh quality reads [are] grouped into UID-families."[257]

190. Furthermore, PCT '213 Appl. (Vogelstein) discloses that "various amplifications are not perfect, so every strand of every original template molecule is not recovered as a UID family."[258] One of ordinary skill in the art would have understood that at least some reads have no second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides. As described *supra*, only a fraction of the library fragments from the original library are used in the bridge amplification step, resulting in some sequence read having no second tagged complementary strand. Thus, one of ordinary skill in the art would have understood that at least some sequence families in PCT '213 Appl. (Vogelstein) are families of unpaired reads having no second tagged complementary strand derived from cfDNA molecules from among the tagged parent polynucleotides.

191. The fact that one of ordinary skill in the art would have expected not all sequence reads to have a complementary strand would motivate the skilled artisan to combine the teachings of PCT '442 Appl. (Bielas) and PCT '213 Appl.

---

[256] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [19].

[257] *Id*. at ¶ [61]; *see also id.* at ¶ [09].

[258] *Id*. at ¶ [66].

(Vogelstein). Although PCT ’213 Appl. (Vogelstein) primarily discloses grouping sequence reads into single-stranded families, PCT ’213 Appl. (Vogelstein) recognizes that “specificity can be even further increased by requiring that each strand of the original double stranded template contain the mutation . . . .”[259] PCT ’213 Appl. (Vogelstein) thus suggests that considering families of paired sequence reads could further reduce errors when reducing redundancy of the reads, and PCT ’442 Appl. (Bielas) teaches how to use families of paired reads to reduce redundancy.

192. Knowing that not all sequencing reads will have counterparts, one of ordinary skill in the art would been motivated to use PCT ’442 Appl. (Bielas)’s approach of sorting sequence reads into families of paired reads where possible, in order to obtain the additional reduction in errors suggested by both PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein). Nonetheless, one of ordinary skill in the art would have continued to use PCT ’213 Appl. (Vogelstein)’s approach of sorting sequence reads into families of unpaired reads for any sequencing reads that do not have a counterpart. For example, when considering the usefulness of the unpaired read approach, PCT ’213 Appl. (Vogelstein) reported reducing sequencing errors

---

[259] *Id*. at ¶ [67]; *see also id.* at ¶¶ [33], [42].

from 15-fold to 70-fold,[260] which is substantial and would have motivated one of ordinary skill in the art to use the unpaired approach in the absence of the paired approach. In this way, PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) together provide dual benefits—further improved error reduction (from paired reads) and maximal use of sequencing data by considerable error reduction (by also considering unpaired reads).

193. At a minimum, using both paired and unpaired reads would have been obvious to try as one of three options for tracking or reducing redundancy—(*i*) paired reads only, (*ii*) unpaired reads only, or (*iii*) both. Using both paired and unpaired would avoid wasting resources (*e.g.*, discarding unpaired reads) and would be expected to have predictable results given that methods of generating consensus sequences from both unpaired and paired reads were taught by the prior art and generally well-known.

194. Thus, the combination of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) discloses sorting sequence reads into families of paired and unpaired reads.

---

[260] *Id*. at ¶¶ [41], [52], [54].

9. Claim 18: "The method of claim 17, wherein the sample is blood, plasma, or serum."

195. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

196. PCT '442 Appl. (Bielas) describes using "circulating mutant DNA in blood" from a patient.[261]

197. To the extent further disclosure is needed, PCT '213 Appl. (Vogelstein) discloses that "[d]etection of donor DNA in the blood of organ transplant patients is an important indicator of graft rejection and detection of fetal DNA in maternal plasma can be used for prenatal diagnosis in a non-invasive fashion."[262]

10. Claim 20: "The method of claim 17, wherein the tagging comprises ligating the molecular barcodes to double-stranded cfDNA molecules."

198. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

199. PCT '442 Appl. (Bielas) discloses that "[t]arget nucleic acid molecules were ligated into a cypher vector library containing previously catalogued dual, double-stranded cyphers."[263]

---

[261] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 23:14; *see also id*. at 25:27–28.

[262] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [03].

[263] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:9–10; *see also id*. at 4:24–30.

200. To the extent further disclosure is needed, PCT ’213 Appl. (Vogelstein) discloses “UIDs, sometimes called barcodes or indexes, can be assigned to nucleic acid fragments in many ways. These include the introduction of exogenous sequences through PCR or ligation.”[264]

11. Claim 21: “The method of claim 17, wherein the set of tags comprises 2 to 10,000 different molecular barcode sequences.”

201. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.

202. PCT ’442 Appl. (Bielas) discloses molecular barcodes “having a length ranging from about 5 to about 50 nucleotides.”[265] For example, PCT ’442 Appl. (Bielas) discloses that the same barcode can be used on each end of the molecule,[266] meaning that barcode length of 5 or 6 nucleotides discloses the claimed range of “2 to 10,000 different molecular barcode sequences” (*i.e.*, $4^5 = 1{,}024$ and $4^6 = 4{,}096$).

203. To the extent further disclosure is needed, PCT ’213 Appl. (Vogelstein)’s disclosure of UIDs with 2 to 4,000 nucleotides, as well as its disclosure of a broad range of number of different unique identifier sequences,

---

[264] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [39] (citations omitted); *see also id*. at ¶ [22].

[265] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 8:19–21.

[266] *Id*. at 3:9–10; *see also id*. at 5:28–30.

comprises the claimed "2 to 10,000 different molecular barcode sequences."[267] For example, UIDs with two (2) nucleotides on both ends of the DNA fragment would provide 256 barcodes. PCT '213 Appl. (Vogelstein) further discloses a range of numbers of different combinations of molecular barcodes that includes "at least 4, at least 16, at least 64, at least 256, at least 1,024, at least 4,096 . . . different sequences."[268]

12. Claim 24: "The method of claim 17, wherein (d) further comprises reducing or tracking redundancy of a plurality of sequence reads from the set of sequencing reads, wherein the reducing or tracking comprises mapping at least a subset of the plurality of sequence reads to a reference sequence, and the reducing or tracking is based on (i) the molecular barcodes associated the paired reads or the unpaired reads and (ii) sequence information at the start and stop positions of the paired reads or the unpaired reads."

204. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

205. As described *supra* regarding Element [1f] of the '306 Patent, PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses

---

[267] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶¶ [22], [30].

[268] *Id*. at ¶ [30].

reducing or tracking the redundancy of sequence reads by mapping at least a subset of the plurality of sequence reads to a reference sequence.[269]

206. As described *supra* regarding Element [1e] and Claim 13 of the ’306 Patent, PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses reducing or tracking the redundancy of sequence reads based on the molecular barcodes associated with the paired reads or the unpaired reads.[270]

207. As described *supra* regarding Claim 13 of the ’306 Patent, PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses reducing or tracking the redundancy of sequence reads based on sequence information at the start and stop positions of the paired reads or the unpaired reads.[271]

13. Claim 25: “The method of claim 24, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises determining a base call at one or more genomic loci for a plurality of sequence reads that map to the one or more genomic loci on the reference sequence.”

208. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.

---

[269] *See supra* at § X.A.2(h).

[270] *See supra* at §§ X.A.2(g) and X.A.7.

[271] *See supra* at § X.A.7.

209. For example, PCT ’442 Appl. (Bielas) explains that “[b]ase calling and sequence alignment will be performed using, for example, the Eland pipeline (Illumina, San Diego, CA).”[272] The base calling and sequence alignment disclosed in PCT ’442 Appl. (Bielas) is a step in which raw sequence reads are mapped to a reference sequence in connection with reducing redundancy by generating a consensus sequence.

210. To the extent further disclosure is required, PCT ’213 Appl. (Vogelstein) discloses that after sequencing, “[b]ase-calling and sequence alignment was performed with the Eland pipeline (Illumina).”[273] PCT ’213 Appl. (Vogelstein) discloses that after mapping the sequence reads “[o]nly high quality reads . . . were used for subsequent analysis.”[274] PCT ’213 Appl. (Vogelstein) then explains that these “[h]igh quality reads were grouped into UID-families based on their endogenous or exogenous UIDs.”[275] Thus, one of ordinary skill in the art would have understood PCT ’213 Appl. (Vogelstein) discloses mapping at least a subset of

---

[272] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 25:9–10; *see also id*. at 11:6–19. I note that one of ordinary skill in the art would have known that “base calling” means determining the most likely nucleotide for that particular base position based on the fluorescent signal detected by the sequencer.

[273] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [60].

[274] *Id*.

[275] *Id*. at ¶ [61].

the plurality of sequence reads to the reference sequence as part of the process of reducing or tracking the redundancy of the plurality of sequence reads.

14. Claim 26: "The method of claim 24, wherein reducing or tracking the redundancy of the plurality of sequence reads comprises collapsing the plurality of sequence reads to produce consensus sequences representative of a sequence of the original double-stranded cfDNA molecules from among the tagged parent polynucleotides."

211. It is my opinion that PCT '442 Appl. (Bielas) alone or in combination with PCT '213 Appl. (Vogelstein) discloses this Claim element.

212. PCT '442 Appl. (Bielas) discloses that "[a]ll sequencing reads having identical cypher pairs, along with their reverse complements, were grouped into families. Comparison of family sequences allowed for generation of a consensus sequence wherein 'mutations' (errors) arising during library preparation (open circle) and during sequencing (gray circle and triangle) were computationally eliminated."[276] This is shown *infra* in Figure 3B of the PCT '442 Appl. (Bielas).[277]

[276] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:11–15.

[277] *Id*. at pp. 40–41.

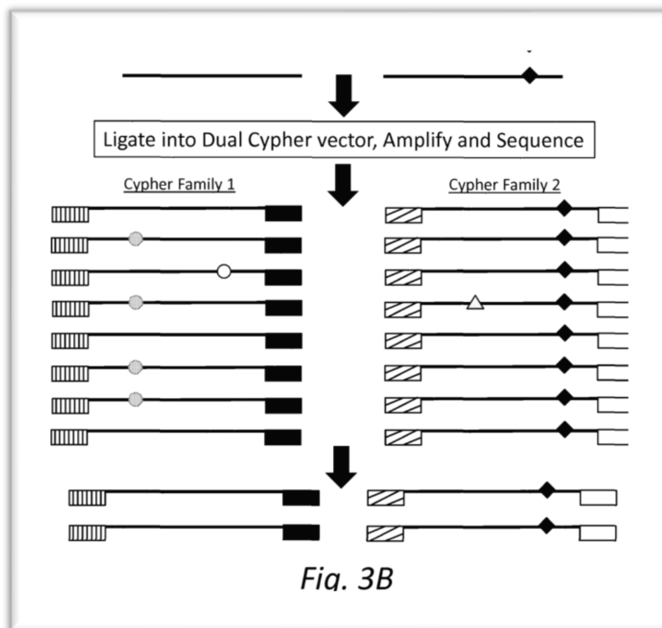


Fig. 3B

213. To the extent further disclosure is required, PCT ’213 Appl. (Vogelstein) discloses that after sequencing, “[h]igh quality reads [are] grouped into UID-families based on their endogenous or exogenous UIDs.”[278] One of ordinary skill in the art would have understood that this is done to generate consensus sequences.

15. Claim 29: “The method of claim 1, wherein the distinct cfDNA molecules in (e) are determined based on (i) the paired reads and (ii) the unpaired reads.”

214. It is my opinion that PCT ’442 Appl. (Bielas) alone or in combination with PCT ’213 Appl. (Vogelstein) discloses this Claim element.

---

[278] *Id*. at ¶ [61].

215. As described *infra* in connection with Element [1e] of the '306 Patent, PCT '442 Appl. (Bielas) discloses determining distinct cfDNA molecules based on paired reads.[279]

216. As described *infra* in connection with Element [1e] of the '306 Patent, PCT '213 Appl. (Vogelstein) discloses determining distinct cfDNA molecules based on unpaired reads.[280]

217. As described *infra* in connection with Element [17d] of the '306 Patent, it would have been obvious in view of PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), and one of ordinary skill in the art would have been motivated, to use PCT '442 Appl. (Bielas)'s approach of reducing redundancy based on families of paired reads, in order to obtain the additional reduction in errors suggested by both PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), but continue to use PCT '213 Appl. (Vogelstein)'s approach of reducing redundancy based on families of unpaired reads for any sequencing reads that do not have a counterpart.[281]

---

[279] *See supra* at § X.A.2(f).

[280] *Id*.

[281] *See supra* at § X.A.8(e).

B. Ground 2: Claims 3, 9, 10, 11, 19 and 22-23 Are Rendered Obvious By PCT '442 Appl. (Bielas) in View of PCT '213 Appl. (Vogelstein) and Forshew (2012)

218. For the reasons set forth *infra*, Challenged Claims 3, 9, 10, 11, 19 and 22-23 of the '306 Patent are invalid as obvious over PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein) and Forshew (2012).

1. Motivation To Combine

219. One of ordinary skill in the art would have been motivated to combine PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein)'s teachings, as described *supra*, with Forshew (2012)'s teaching of using 0.9 ng to 19.7 ng of cfDNA molecules.[282]

220. Forshew (2012) is analogous art to PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), in that each of these references is directed to methods for identifying genetic mutations in DNA through next generation sequencing using barcodes. Furthermore, both Forshew (2012) and PCT '442 Appl. (Bielas) disclose sequencing "[c]irculating cell-free DNA extracted from plasma or other body fluids."[283] One of ordinary skill in the art would have been motivated, in

[282] *See* Ex. 1010 (Forshew (2012)) at Table 2.

[283] *Id*. at p. 1; *see also* Ex. 1005 (PCT '442 Appl. (Bielas)) at 23:12–14; *see also id*. at 25:27–28.

implementing PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein)’s teachings, to optimize the amount of cfDNA for amplification and sequencing, particularly because the concentration of cfDNA is expected to be relatively low in natural samples.[284] One of ordinary skill in the art would have been motivated to look to Forshew (2012)’s examples of amounts of cfDNA molecules contained in different sample volumes sufficient for use in identifying mutations in cfDNA.

221. One of ordinary skill in the art also would have been motivated to combine PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein)’s teachings with Forshew (2012)’s teaching of selective enrichment and examples of cancer genes. Indeed, Forshew (2012) used a preamplification step employing multiplex PCR followed by singleplex PCR to amplified 11 target regions.[285] Both PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) disclose methods for identifying mutations present in small fractions of DNA templates.[286] One of ordinary skill in the art would have been motivated, in implementing PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein)’s teachings, to enrich specific DNA sequences of known cancer-associated genes for more efficient sequencing. Indeed, PCT ’213

---

[284] *See* Ex. 1010 (Forshew (2012)) at p. 1.

[285] *Id.* at p. 3; Figure 1.

[286] *See*, *e.g.*, Ex. 1005 (PCT ’442 Appl. (Bielas)) at 1:21–2:6; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at Abstract.

Appl. (Vogelstein) discloses the use of a hybridization capture step to target genes of interest.[287] Enrichment techniques are particularly important when using cfDNA, which is present in relatively small amounts.[288]

222. One of ordinary skill in the art would also have been motivated to combine PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) with Forshew (2012), as PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) disclose that their methods are useful for cancer diagnostics, and Forshew (2012) pinpoints certain (known) genes of interest such as TP53 and EGFR for that same purpose.[289]

223. One of ordinary skill in the art combining PCT ’442 Appl. (Bielas), PCT ’213 Appl. (Vogelstein), and Forshew (2012) would have had a reasonable expectation of success in arriving at the methods of Claims 3, 9-11, 19, and 22-23 of the ’306 Patent. For example, the prior art recognizes that cfDNA may be present in relatively low concentrations, as it teaches modifications to protocols for lower amounts of DNA, such as PCT ’442 Appl. (Bielas)’s “amplify[ing] the number of each molecule” through PCR and digital PCR and PCT ’213 Appl. (Vogelstein)’s

---

[287] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [25].

[288] *See* Ex. 1010 (Forshew (2012)) at p. 1.

[289] *Id*. at p. 3.

increased number of PCR cycles to address low amounts of DNA molecules.[290] The art also disclosed numerous commercial kits for extracting DNA from blood.[291] One of ordinary skill in the art would have recognized that Forshew (2012)'s teaching of 0.9 to 19.7 ng of cfDNA molecules could be used as a starting point for PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein)'s methods.

224. One of ordinary skill in the art would have further expected that PCT '213 Appl. (Vogelstein)'s enrichment techniques could be used on the well-known cancer genes noted in Forshew (2012). One of ordinary skill in the art would have expected to successfully perform Forshew (2012)'s second round of singleplex amplification to enrich target DNA because PCR amplification was a routine technique known in the field.

2. Claim 3: "The method of claim 1, wherein the population of cfDNA molecules comprises 1 nanogram (ng) to 100 ng of cfDNA molecules."

225. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.

---

[290] *See*, *e.g.*, Ex. 1005 (PCT '442 Appl. (Bielas)) at 1:21–2:6; *see also* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at Abstract.

[291] *See*, *e.g.*, Ex. 1005 (PCT '442 Appl. (Bielas)) at p. 28; *see also* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [57].

226. Specifically, Forshew (2012) extracted a range of amounts of cfDNA from plasma samples for sequence library preparation, including between 0.9 and 19.7 ng.[292] One of ordinary skill in the art would have understood that Forshew (2012)'s range of sample amounts, which falls within the claimed range of 1 to 100 ng, could be used with the methods disclosed in PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein).

3. Claim 9: "The method of claim 1, further comprising, prior to the sequencing, enriching at least a subset of the amplified progeny polynucleotides for target regions of interest to produce enriched progeny polynucleotides."

227. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.

228. PCT '213 Appl. (Vogelstein) discloses "[c]omplexity of the analyte fragments can be decreased by a capture step" that will "[t]ypically . . . employ hybridization to probes representing a gene or set of genes of interest."[293] As described *supra*, hybridization capture is a target enrichment technique used to enrich specific DNA sequences in regions of interest.[294] PCT '213 Appl. (Vogelstein) provides an example where the sequences of six genes of interest were

---

[292] *See* Ex. 1010 (Forshew (2012)) at pp.17–18, Table 2, Table S6.

[293] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [25].

[294] *See supra* at § VII.F.3.

selectively enriched that were associated with cancer.[295] One of ordinary skill in the art would have understood that PCT '213 Appl. (Vogelstein)'s "capture step" employing "hybridization" discloses this Claim element.

229. To the extent further disclosure is needed, Forshew (2012) discloses enrichment of amplified progeny polynucleotides for target regions of interest. For example, Forshew (2012) discloses amplifying the "selected regions" of certain cancer-associated genes.[296] Forshew (2012) explains that the target regions are enriched by "amplif[ying] DNA from each sample in duplicate . . . producing mean read depth of 3200 above Q30 for each of the 9024 expected read groups (48 amplicons x 2 directions x 94 barcoded samples)."[297] One of ordinary skill in the art would have understood that a preliminary round of amplification can be useful to generate "amplified progeny polynucleotides" as a target enrichment step. One of ordinary skill in the art would have understood that Forshew (2012)'s second round of singleplex amplification of "selected regions" "in duplicate" and at high accuracy (Q30) increases the proportion of those selected regions within the sample.

---

[295] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [40].

[296] *See* Ex. 1010 (Forshew (2012)) at p.3 (explaining that primer pairs targeted exons and exon junctions in the following genes: *TP53*, *PTEN*, *EGFR*, *BRAF*, *KRAS*, *PIK3CA*).

[297] *Id.*

4. Claim 10: "The method of claim 9, wherein the target regions of interest comprise genetic sequences of a plurality of genes selected from the group consisting of ALK, APC, BRAF, CDKN2A, EGFR, ERBB2, FBXW7, KRAS, MYC, NOTCH1, NRAS, PIK3CA, PTEN, RB1, TP53, MET, AR, ABL1, AKT1, ATM, CDH1, CSF1R, CTNNB1, ERBB4, EZH2, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAQ, GNAS, HNF1A, HRAS, IDH1, IDH2, JAK2, JAK3, KDR, KIT, MLH1, MPL, NPM1, PDGFRA, PROC, PTPN11, RET, SMAD4, SMARCB1, SMO, SRC, STK11, VHL, TERT, CCND1, CDK4, CDKN2B, RAF 1, BRCA1, CCND2, CDK6, NF1, TP53, ARID1A, BRCA2, CCNE1, ESR1, RIT1, GATA3, MAP2K1, RHEB, ROS1, ARAF, MAP2K2, NFE2L2, RHOA, and NTRK1."

230. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.

231. Forshew (2012) discloses designing primer pairs "to amplify 5995 bases of genomic sequence covering coding regions (exons and exon junctions) of *TP53* and *PTEN*, and selected regions in *EGFR*, *BRAF*, *KRAS*, and *PIK3CA* (table S1) by overlapping short amplicons (Fig. 1A)."[298] One of ordinary skill in the art would have understood that Forshew (2012)'s targeting of regions in *TP53*, *EGFR*, *BRAF*, and *KRAS* disclose or render obvious the claimed "plurality of genes selected from the group . . ." because the claimed genes are cancer genes expressly disclosed and analyzed by Forshew (2012).

---

[298] *Id.*

5. Claim 11: "The method of claim 10, further comprising, prior to the sequencing, amplifying a plurality of the enriched progeny polynucleotides."

232. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.

233. For example, PCT '213 Appl. (Vogelstein) discloses the sequential steps of enrichment, amplification, and sequencing.[299] For example, PCT '213 Appl. (Vogelstein) describes an experiment where a first round of "[a]mplification of a fragment of human genomic DNA" was performed, resulting in enriched progeny polynucleotides; then, a second round of amplification of the enriched progeny polynucleotides occurred using "[f]ifty uL PCR reactions."[300] PCT '213 Appl. (Vogelstein) then explains that "[t]he entire contents of each well were then used as templates for the exogenous UIDs strategy described above," including sequencing.[301] One of ordinary skill in the art would have understood that PCT '213 Appl. (Vogelstein) discloses Claim 11 of the '306 Patent because PCT '213 Appl. (Vogelstein)'s "Post-Capture Amplification" discloses post-enrichment amplification of enriched polynucleotides.[302]

---

[299] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶¶ [09], [13], [55].

[300] *Id.* at ¶ [59].

[301] *Id.*

[302] *Id.* at ¶ [57], Table 6.

234. To the extent further disclosure is needed, PCT '442 Appl. (Bielas) discloses amplifying a plurality of progeny polynucleotides prior to sequencing.[303] To the extent one of ordinary skill in the art would have enriched at least a subset of the amplified progeny polynucleotides for target regions of interest using the methods of PCT '213 Appl. (Vogelstein) or Forshew (2012) with the methods disclosed by PCT '442 Appl. (Bielas), it would have been obvious to amplify a plurality of enriched progeny polynucleotides prior to sequencing.

6. Claim 19: "The method of claim 17, wherein the population of double-stranded cfDNA molecules comprises 1 nanogram (ng) to 100 ng of double-stranded cfDNA molecules."

235. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.

236. For example, PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and Forshew (2012) discloses using 1 ng to 100 ng of double-stranded cfDNA molecules. Specifically, Forshew (2012) extracted a range of amounts of cfDNA from plasma samples for sequence library preparation, including between 0.9 and 19.7 ng.[304] One of ordinary skill in the art would have understood that Forshew (2012)'s range of sample amounts, which falls within the claimed

---

[303] *See supra* at § X.A.2(e) regarding Claim element [1d] of the '306 Patent.

[304] *See* Ex. 1010 (Forshew (2012)) at pp.17–18, Table 2, Table 6.

range of 1 to 100 ng, could be used with the methods disclosed in PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein).

7. Claim 22: “The method of claim 17, further comprising, prior to the sequencing, enriching at least a subset of the amplified progeny polynucleotides for target regions of interest to produce enriched progeny polynucleotides.”

237. It is my opinion that PCT ’442 Appl. (Bielas) in combination with PCT ’213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.[305]

8. Claim 23: “The method of claim 22, wherein the target regions of interest comprise genetic sequences of a plurality of genes selected from the group consisting of ALK, APC, BRAF, CDKN2A, EGFR, ERBB2, FBXW7, KRAS, MYC, NOTCH1, NRAS, PIK3CA, PTEN, RB1, TP53, MET, AR, ABL1, AKT1, ATM, CDH1, CSF1R, CTNNB1, ERBB4, EZH2, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAQ, GNAS, HNF1A, HRAS, IDH1, IDH2, JAK2, JAK3, KDR, KIT, MLH1, MPL, NPM1, PDGFRA, PROC, PTPN11, RET, SMAD4, SMARCB1, SMO, SRC, STK11, VHL, TERT, CCND1, CDK4, CDKN2B, RAF1, BRCA1, CCND2, CDK6, NF1, TP53, ARID1A, BRCA2, CCNE1, ESR1, RIT1, GATA3, MAP2K1, RHEB, ROS1, ARAF, MAP2K2, NFE2L2, RHOA, and NTRK1.”

238. It is my opinion that PCT ’442 Appl. (Bielas) in combination with PCT ’213 Appl. (Vogelstein) and Forshew (2012) discloses this Claim element.[306]

---

[305] *See supra* at § X.B.3 regarding Claim 9 of the ’306 Patent.

[306] *See supra* at § X.B.4 regarding Claim 10 of the ’306 Patent.

C. Ground 3: Claims 4, 5 and 6 Are Rendered Obvious By PCT ’442 Appl. (Bielas) in View of PCT ’213 Appl. (Vogelstein), PCT ’832 Appl. (Hendricks) and/or ’081 Patent Appl. (Diehn)

239. For the reasons set forth *infra*, Challenged Claims 4, 5 and 6 of the ’306 Patent are invalid as obvious over PCT ’442 Appl. (Bielas) in view of PCT ’213 Appl. (Vogelstein), PCT ’832 Appl. (Hendricks) and/or the ’081 Patent Appl. (Diehn).

1. Motivation to Combine

240. One of ordinary skill in the art would have been motivated to combine PCT ’442 Appl. (Bielas)’s and PCT ’213 Appl. (Vogelstein)’s teachings, as described *supra*, with PCT ’832 Appl. (Hendricks)’s teaching of ligation efficiency and molar excess, and the ’081 Patent Appl. (Diehn)’s teaching of molar excess.

241. Like PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein), PCT ’832 Appl. (Hendricks) is analogous art that is directed to methods used for nucleic acid sequencing.[307] Both PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) disclose ligation as a preferred means of attaching barcodes to target nucleic acid molecules.[308] PCT ’213 Appl. (Vogelstein) teaches that the barcodes should be in

[307] *See*, *e.g.*, Ex. 1008 (PCT ’832 Appl. (Hendricks)) at [0010], [0014], [0073], [0079], [0092].

[308] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 4:9–10; Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [39].

excess of the target DNA molecules.[309] One of ordinary skill in the art would have been motivated, in implementing PCT '442 Appl. (Bielas)'s and PCT '213 Appl. (Vogelstein)'s teachings, to look to PCT '832 Appl. (Hendricks)'s ligation methods to "produce a significant or detectable amount of ligation product" in order to maximize the sequenceable target DNA, and thus ensure high quality sequencing data, especially when starting with relatively low amounts of cfDNA.[310]

242. One of ordinary skill in the art would also have been motivated to combine PCT '442 Appl. (Bielas), PCT '213 Appl. (Vogelstein), and PCT '832 Appl. (Hendricks) with the '081 Patent Appl. (Diehn). Like PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), the '081 Patent Appl. (Diehn) is analogous art directed to methods used for identifying genetic mutations using nucleic acid sequencing.[311] Furthermore, both the '081 Patent Appl. (Diehn) and PCT '442 Appl. (Bielas) disclose sequencing cfDNA.[312] The '081 Patent Appl. (Diehn) explains that "[i]ncreasing adapter concentration during ligation increases ligation efficiency and

---

[309] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [08].

[310] *See* Ex. 1008 (PCT '832 Appl. (Hendricks)) at ¶ [0076].

[311] *See*, *e.g.*, Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶¶ [0027], [0044], [0050], [0065], [0068], [00798], [0082], [0093], [0104], [0107].

[312] *See* Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶¶ [0044], [0050], [0082], [0093]; *see also* Ex. 1005 (PCT '442 Appl. (Bielas)) at 23:12–14, 25:27–28.

reporter recovery."[313] One of ordinary skill in the art would have known this direct correlation between increasing concentration results in increased ligation efficiency and thus been motivated by the '081 Patent Appl. (Diehn)'s use of molar excess to obtain the range of ligation efficiencies described in PCT '213 Appl. (Hendricks).[314]

243. One of ordinary skill in the art combining PCT '442 Appl. (Bielas), PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks), and/or the '081 Patent Appl. (Diehn) would have had a reasonable expectation of success in arriving at the methods of Claims 4-6 of the '306 Patent. One of ordinary skill in the art would have expected that using a higher molar excess of tags relative to DNA molecules would achieve a higher ligation efficiency, thus increasing the number of tagged molecules and thus improving overall effectiveness of the downstream steps.[315] Indeed, the '081 Patent Appl. (Diehn) expressly notes the direct correlation of "***[i]ncreasing adapter concentration during ligation increases ligation efficiency*** and reporter recovery."[316] One of ordinary skill in the art would also have

---

[313] *See* Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶ [0130].

[314] *Id*. at ¶¶ [0130], [0157]; *see also* Ex. 1008 (PCT '832 Appl. (Hendricks)) at ¶ [0076].

[315] *See* Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶ [0157].

[316] *Id.* at ¶ [0130] (emphasis added).

appreciated that increasing the adapter-DNA ligation efficiency was beneficial when using relatively low amounts of DNA, such as cfDNA and more particularly ctDNA.

2. Claim 4: "The method of claim 1, wherein the molecular barcodes are ligated to the plurality of the cfDNA molecules using more than a 10× excess of duplex tags as compared to the population of cfDNA molecules, wherein at least 20% of the cfDNA molecules from the population are tagged with the duplex tags."

244. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks) and/or '081 Patent Appl. (Diehn) discloses Claim 4 of the '306 Patent.

245. For example, PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein) disclose ligation of barcodes to target nucleic acid molecules.[317] PCT '832 Appl. (Hendricks) discloses a ligation process that, if used with the teachings of PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), would result in at least 20% or at least 40% of the cfDNA molecules being tagged with the duplex tags. Specifically, PCT '832 Appl. (Hendricks) discloses ligation that "produce[s] a significant or detectable amount of ligation product. Optionally, the efficiency of ligation is at least . . . 20%, 30%, 50%, 60%, 70%, 75%, 80% . . .."[318] PCT '832

[317] *See* Ex. 1005 (PCT '442 Appl. (Bielas)) at 4:9–10; *see also* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [39].

[318] *See* Ex. 1008 (PCT '832 Appl. (Hendricks)) at ¶ [0076].

Appl. (Hendricks)'s range of ligation efficiency of "at least . . . 20%" or "at least . . . 50%," in addition to its express disclosure of higher efficiencies, discloses this element.

246. PCT '832 Appl. (Hendricks) also discloses ligating adaptors using more than an 10× excess and more than an 80× excess of tags as compared to DNA molecules. PCT '832 Appl. (Hendricks) discloses that "[l]igation reactions were performed in a 96 well plate using 2 nM template/primer, 2.0 μM ligase, 2-5 μM short oligonucleotide in a total volume of 50 μL."[319] One of ordinary skill in the art would have understood that using the lower end of the range of the short oligonucleotide of (2 μM) calculations comparing the 2 nM template/primer concentration would show a 1000× molar excess (*i.e.*, 2 μM = 2,000 nM; whereby 2,000 nM / 2 nM = 1000× molar excess), and using the higher end range of concentration (5 μM) results in a 2500× molar excess (*i.e.*, 5 μM = 5,000 nM; whereby 5,000 nM / 2 nM = 2500× molar excess). Thus, PCT '832 Appl. (Hendricks) discloses the claimed "more than a 10× excess" and "more than an 80× excess."

247. To the extent additional disclosure is needed, the '081 Patent Appl. (Diehn) discloses ligating adaptors to cfDNA molecules using more than an 10×

---

[319] *Id.* at ¶ [00138].

excess or more than an 80× excess of tags as compared to molecules. The '081 Patent Appl. (Diehn) discloses starting with "[c]ell-free DNA (cfDNA) . . . isolated from 1-5 mL plasma with the QIAamp Circulating Nucleic Acid Kit (Qiagen)."[320] Prior to ligation, that DNA was added to "2 concentrations of Illumina adapters in the ligation reaction: 12 nM (10-fold molar excess to cfDNA fragments) and 120 nM (100-fold molar excess)."[321] One of ordinary skill in the art would have understood that the adapter ligation process disclosed in the '081 Patent Appl. (Diehn) would be applicable to PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), and that the '081 Patent Appl. (Diehn)'s 100-fold molar excess (*i.e.*, 100× molar excess) discloses the claimed ranges of "more than a 10× excess" and more than an 80× excess."

3. Claim 5: "The method of claim 1, wherein the molecular barcodes are ligated to the plurality of the cfDNA molecules using more than an 80× excess of duplex tags as compared to the population of cfDNA molecules."

248. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks) and/or the '081 Patent Appl. (Diehn) discloses Claim 5 of the '306 Patent.[322]

---

[320] *See* Ex. 1009 ('081 Patent Appl. (Diehn)) at ¶ [0141]; *see also id.* at ¶ [0042].

[321] *Id.* at ¶ [0159]; *see also id.* at ¶ [0143].

[322] *See supra* at § X.C.2 regarding Claim 4 of the '306 Patent.

4. Claim 6: "The method of claim 5, wherein at least 40% of the cfDNA molecules from the population are tagged with the duplex tags."

249. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein), PCT '832 Appl. (Hendricks) and/or the '081 Patent Appl. (Diehn) discloses Claim 6 of the '306 Patent.[323]

D. Ground 4: Claims 14-16 and 27-28 Are Rendered Obvious By PCT '442 Appl. (Bielas) in View of PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks)

250. For the reasons set forth *infra*, Challenged Claims 14-16 and 27-28 of the '306 Patent are invalid as obvious over PCT '442 Appl. (Bielas) in view of PCT '213 Appl. (Vogelstein) and '156 Patent (Hicks).

1. Motivation to Combine

251. In my opinion, one of ordinary skill in the art would have been motivated to combine the teachings of PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein), as described *supra*, with the teaching of the '156 Patent (Hicks) concerning quantitatively measuring sequencing reads and estimating tagged parent polynucleotides.

252. The '156 Patent (Hicks) is analogous art to PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein) as it is directed to methods used for nucleic acid

---

[323] *See supra* at § X.C.2 regarding Claim 4 of the '306 Patent.

sequencing.[324] The methods of accurately detecting mutations in nucleic acid sequences disclosed in PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein)[325] are important to enable determination or estimation of the frequency of mutations, and in particular rare genetic mutations or variants. This is recognized, for example, in PCT ’213 Appl. (Vogelstein)’s example where “[t]he exogenous UID strategy (Fig. 3) was then used to determine the prevalence of rare mutations . . . .”[326] One of ordinary skill in the art would have been motivated, in implementing the teachings of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein), to use the methods disclosed in the ’156 Patent (Hicks) to estimate the number of cfDNA molecules that map to particular loci because such data would be useful for calculating copy number variations, which in turn is useful for the sensitive diagnosis of genetic conditions.

253. One of ordinary skill in the art combining PCT ’442 Appl. (Bielas), PCT ’213 Appl. (Vogelstein), and the ’156 Patent (Hicks) would have had a reasonable expectation of success in arriving at the methods of Claims 14-16 and 27-28 of the ’306 Patent. Like PCT ’442 Appl. (Bielas) and PCT ’213 Appl.

---

[324] *See*, *e.g.*, Ex. 1011 (’156 Patent (Hicks)) at Abstract.

[325] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 3:17–22; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at Abstract.

[326] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [52]; *see also id.* ¶ [31]; *see also* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 1:21–2:27.

(Vogelstein), the ’156 Patent (Hicks) discloses using barcodes to track original target molecules.[327] One of ordinary skill in the art would have expected the tagging methods in PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) could be used with the quantification methods in the ’156 Patent (Hicks) because in each reference the barcodes are used to link sequence reads back to original molecules.[328] Each reference the barcodes is formed by attaching barcodes to target molecules using known techniques.[329]

2. Claim 14: “The method of claim 1, further comprising: (f) determining quantitative measures of (i) paired reads or (ii) unpaired reads that map to a genomic locus of the reference sequence.”

254. It is my opinion that PCT ’442 Appl. (Bielas) in combination with PCT ’213 Appl. (Vogelstein) and the ’156 Patent (Hicks) discloses Claim 14 of the ’306 Patent.

255. As described previously,[330] PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) disclose mapping at least a subset of the plurality of sequence

---

[327] *See, e.g.*, Ex. 1011 (’156 Patent (Hicks)) at 19:11–17.

[328] *See* Ex. 1005 (PCT ’442 Appl. (Bielas)) at 4:24; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [19]; Ex. 1011 (’156 Patent (Hicks)) at 19:11–27.

[329] *See, e.g.*, Ex. 1005 (PCT ’442 Appl. (Bielas)) at 9:22–26; *see also* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [22]; Ex. 1011 (’156 Patent (Hicks)) at 18:5–8.

[330] *See supra* at § X.A.2(h).

reads to a reference sequence. PCT ’213 Appl. (Vogelstein) also discloses that its methods “can be used to quantitate,” including to compare “relative abundance of two analyte DNA fragments.”[331] For example, PCT ’213 Appl. (Vogelstein) discloses examples where the number of families are quantified, as well as the number of mutations represented by the families.[332]

256. To the extent additional disclosure is required, the ’156 Patent (Hicks) discloses determining quantitative measures of sequence reads that map to a reference sequence. The ’156 Patent (Hicks) explains that “[a]fter deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted.”[333] One of ordinary skill in the art would have understood that “deconvoluting by sample tags” reduces each family of polynucleotides to a single sequence, and the number of sequences is then counted to determine the number of tagged nucleic acid molecules mapped to each location.

257. If one of ordinary skill in the art used the teachings of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein) to generate families of paired and unpaired reads, as described *supra* for Claims 1 and 17 of the ’306 Patent, it would

---

[331] *See* Ex. 1006 (PCT ’213 Appl. (Vogelstein)) at ¶ [31].

[332] *Id*. at ¶ [37]; *see also id.* at ¶ [61].

[333] *See* Ex. 1011 (’156 Patent (Hicks)) at 19:36–39; *see also id.* 18:10–13.

have been obvious to quantify the number of paired or unpaired reads that map to a genomic locus of the reference sequence using the techniques described in PCT ’213 Appl. (Vogelstein) and the ’156 Patent (Hicks).

3. Claim 15: “The method of claim 14, further comprising: (g) estimating with a programmed computer processor a quantitative measure of tagged parent polynucleotides that map to the genomic locus based on the quantitative measures of the paired reads and the unpaired reads.”

258. It is my opinion that PCT ’442 Appl. (Bielas) in combination with PCT ’213 Appl. (Vogelstein) and the ’156 Patent (Hicks) discloses this Claim element.

259. The ’156 Patent (Hicks) discloses a quantitative measure of tagged parent polynucleotides can be estimated using quantitative measures of the number of families that map to a genomic locus. Specifically, as described *supra*, the ’156 Patent (Hicks) explains that “[a]fter deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted.”[334] For example, the ’156 Patent (Hicks) discloses a counting method.

> Therefore, for a given region comprising multiple locations, the maximum copy number of the region is not less than the maximum number of different tagged nucleic acid molecules mapped to any location in that region. ***The maximum number of different tagged nucleic acid molecules centered about a moving window of a fixed number of locations can be taken as the measurement of true copy number of the window***. This will provide at worst an underestimate of

[334] *See* Ex. 1011 (’156 Patent (Hicks)) at 19:36–39; *see also id.* 18:10–13.

> the maximum true copy number for that window, and never an overestimate.[335]

260. Thus, one of ordinary skill in the art would have understood that in the ’156 Patent (Hicks), the number of different families mapped to a location are used to count the tagged parent polynucleotides that map to the genomic locus. It would have been obvious to one of ordinary skill in the art that the quantification method in the ’156 Patent (Hicks) could be used with the quantified families of paired and unpaired reads formed by the combined methods of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein).

261. Furthermore, one of ordinary skill in the art would have understood that the calculations performed in the ’156 Patent (Hicks) would be performed using a programmed computer processor.[336] One of ordinary skill in the art would have understood that several of the Figures provided in the ’156 Patent (Hicks) depicting its methods of quantification were generated using a computer processor.[337] As described *supra*,[338] use of computer processors to perform such genome analysis of NGS read data was widely used and routine by December 2013.

---

[335] *Id*. at 19:44–53 (emphasis added).

[336] *See supra* at § VII.I.

[337] *See* Ex. 1011 (’156 Patent (Hicks)) at Figs. 2–4, 6.

[338] *See supra* at § VII.I.

4. Claim 16: "The method of claim 15, further comprising: (h) detecting copy number variation in the population of cfDNA molecules by determining a normalized quantitative measure determined in (g) at each of one or more genomic loci."

262. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks) discloses this Claim element.

263. The '156 Patent (Hicks) discloses that the reference genome is "artificially divide[d] . . . into 'bins' of several kilobases . . . . Each bin contains a number of random genomic sequence reads from which ***a mean value is computed*** to determine genomic copy number at low sequence coverage."[339] One of ordinary skill in the art would have understood that the '156 Patent (Hicks)'s method using bins to calculate a mean value to determine genomic copy number discloses the claimed "detecting copy number variation in the population of cfDNA molecules by determining a normalized quantitative measure determined in (g) at each of one or more genomic loci."

5. Claim 27: "The method of claim 24, further comprising determining quantitative measures of at least two of (i) paired reads that map to one or more genomic loci, (ii) unpaired reads that map to one or more genomic loci, (iii) a

[339] *See* Ex. 1011 ('156 Patent (Hicks)) at 18:15–22 (emphasis added).

read depth of the paired reads, and (iv) a read depth of the unpaired reads."

264. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks) discloses this Claim element. Specifically, PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks) discloses determining quantitative measures of paired reads and unpaired reads that map to a genomic locus of the reference sequence.

265. As described previously,[340] PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein) disclose mapping at least a subset of the plurality of sequence reads to a reference sequence. PCT '213 Appl. (Vogelstein) also discloses that its methods "can be used to quantitate," including to compare "relative abundance of two analyte DNA fragments."[341] For example, PCT '213 Appl. (Vogelstein) discloses examples where the number of families are quantified, as well as the number of mutations represented by the families.[342]

266. To the extent additional disclosure is required, the '156 Patent (Hicks) discloses determining quantitative measures of sequence reads that map to a

[340] *See supra* at § X.A.2(h).

[341] *See* Ex. 1006 (PCT '213 Appl. (Vogelstein)) at ¶ [31].

[342] *Id.* at ¶ [37]; *see also id.* at ¶ [61].

reference sequence. The '156 Patent (Hicks) explains that "[a]fter deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted."[343] One of ordinary skill in the art would have understood that "deconvoluting by sample tags" reduces each family of polynucleotides to a single sequence, and the number of sequences is then counted to determine the number of tagged nucleic acid molecules mapped to each location.

267. If one of ordinary skill in the art used the teachings of PCT '442 Appl. (Bielas) and PCT '213 Appl. (Vogelstein) to generate families of paired and unpaired reads, as described *supra* for Claims 1 and 17 of the '306 Patent, it would have been obvious to quantify the number of paired and unpaired reads that map to a genomic locus of the reference sequence using the described techniques in PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks).

6. Claim 28: "The method of claim 27, further comprising estimating with a programmed computer processor a quantitative measure of tagged parent polynucleotides that map to a genomic locus of the reference sequence based on the quantitative measures of at least two of (i) paired reads,

[343] Ex. 1011 ('156 Patent (Hicks)) at 19:36–39; *see also id.* at 18:10–13.

(ii) unpaired reads, (iii) a read depth of the paired reads, and (iv) a read depth of the unpaired reads at the genomic locus."

268. It is my opinion that PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks) discloses this Claim element. Specifically, PCT '442 Appl. (Bielas) in combination with PCT '213 Appl. (Vogelstein) and the '156 Patent (Hicks) discloses estimating with a programmed computer processor a quantitative measure of tagged parent polynucleotides that map to a genomic locus of the reference sequence based on the quantitative measures of paired reads and unpaired reads.

269. The '156 Patent (Hicks) discloses a quantitative measure of tagged parent polynucleotides can be estimated using quantitative measures of the number of families that map to a genomic locus. Specifically, as described *supra*, the '156 Patent (Hicks) explains that "[a]fter deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted."[344] For example, the '156 Patent (Hicks) discloses a counting method.

> Therefore, for a given region comprising multiple locations, the maximum copy number of the region is not less than the maximum number of different tagged nucleic acid molecules mapped to any location in that region. The maximum number of different tagged nucleic acid molecules centered about a moving window of a fixed number of locations can be taken as the measurement of true copy number of the window. This will provide at worst an underestimate of

[344] *See* Ex. 1011 ('156 Patent (Hicks)) at 19:36–39; *see also id.* at 18:10–13.

> the maximum true copy number for that window, and never an overestimate.[345]

270. Thus, one of ordinary skill in the art would have understood that in the ’156 Patent (Hicks), the number of different families mapped to a location are used to estimate the tagged parent polynucleotides that map to the genomic locus. It would have been obvious to one of ordinary skill in the art that the quantification method in the ’156 Patent (Hicks) could be used with the quantified families of paired and unpaired reads formed by the combined methods of PCT ’442 Appl. (Bielas) and PCT ’213 Appl. (Vogelstein).

271. Furthermore, one of ordinary skill in the art would have understood that the calculations performed in the ’156 Patent (Hicks) would be performed using a programmed computer processor. One of ordinary skill in the art would have understood that several of the Figures provided in the ’156 Patent (Hicks) depicting its methods of quantification were generated using a computer processor.[346] As described *supra*,[347] use of computer processors to perform such genome analysis of NGS read data was widely used and routine by December 2013.

---

[345] *Id*. at 19:44–53 (emphasis added).

[346] Ex. 1011 (’156 Patent (Hicks)) at Figs. 2-4, 6.

[347] *See supra* at § VII.I.

XI. SECONDARY CONSIDERATIONS CANNOT REBUT THE STRONG SHOWING OF OBVIOUSNESS

272. I have been informed by counsel and understand that long-felt but unsolved needs, failure of others, unexpected results, commercial success, copying, licensing, and praise may constitute secondary considerations of non-obviousness.

273. It is my opinion that there are no secondary considerations supporting non-obviousness here for the reasons described *supra*.

I confirm that the contents of this Declaration are true to the best of my knowledge and belief insofar as it states facts and that it contains my honest opinions on the matters upon which I have been asked to give them.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Dated: August 15, 2025

Michael Metzker, Ph.D.