

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

INTEGRATED DNA TECHNOLOGIES, INC.,
Petitioner,

v.

TECAN GENOMICS, INC.,
Patent Owner

Case IPR2025-00015
Case IPR2025-00016
U.S. Patent Nos. 10,036,012 and 10,876,108

DECLARATION OF CARLOS D. BUSTAMANTE, PH.D.

TABLE OF CONTENTS

I.	BACKGROUND AND QUALIFICATIONS	9
II.	MATERIALS CONSIDERED	14
III.	LEGAL STANDARDS	15
IV.	TECHNICAL BACKGROUND	18
A.	Genetic Information	18
B.	Nucleic Acid Genotyping	22
C.	Library Preparation.....	25
D.	Anchored PCR and Genome Walking.....	29
E.	PCR-Free Sequencing Methods	38
F.	Hybridization Arrays	39
G.	Nucleic Acid Sequencing	42
1.	Pre-NGS Technologies for Sequencing.....	42
2.	Next Generation Sequencing.....	45
H.	Target Enrichment	59
1.	PCR-based Target Enrichment Methods	61
2.	Capture-by-Hybridization	66
3.	Capture-by-Circularization.....	68
V.	THE PERSON OF ORDINARY SKILL IN THE ART	69
VI.	THE PATENTED INVENTIONS	70
A.	The Claims of the '012 Patent.....	71
B.	The Claims of the '108 Patent.....	75
C.	Illustration of the Claimed Inventions.....	78
D.	Relevant Prosecution History	87
E.	Claim Construction.....	92
1.	“enrich”	93
2.	“RNA fragment”	98
VII.	RELEVANT PRIOR ART	100
A.	Meyer (EX1006).....	100
1.	The Specific Issues Addressed by Meyer.....	102

2. cDNA Library Preparation	105
3. Sequencing the cDNA Library	117
4. Data Analysis.....	118
B. Siebert (EX1007).....	121
C. Caruccio (EX1008).....	124
D. Bronner (EX1009).....	125
<u>VIII. MEYER FAILS TO DISCLOSE THE CLAIMED INVENTIONS</u>	<u>126</u>
A. “amplifying the one or more oligonucleotide extension products . . . to enrich for the nucleic acid sequence of interest” (’012 patent) / “amplifying the one or more oligonucleotide extension products . . . thereby enriching the nucleic acid sequence of interest” (’108 patent)	126
1. The Disclosed Protocols Do Not Selectively Amplify for cDNA Fragments with Sequences of Interest.....	127
2. Any Enrichment Disclosed by Meyer Occurred Prior to Fragmentation .	136
3. Meyer Does Not Propose Targeted Enrichment of the Broken-T Chain. .	140
B. The Purpose of Meyer Reinforces that it Does Not Disclose Enrichment	141
C. The Data in Meyer Does Not Support IDT’s Claims.....	143
1. A POSA Would Have Understood that the Average Sequencing Read Length Used to Calculate the 10% Estimate Was Skewed.....	148
2. A POSA Would Have Understood that the Reported Transcript Lengths Were Likely Skewed.....	151
3. If Dr. Sims’ Inferences about the Average Transcript Size are Accurate, then a POSA Would Have Concluded that the Estimated Number of 3’ End Reads Should Have Been Approximately 16%, Not 10%.....	155
D. “wherein the nucleic acid fragment comprises a nucleic acid sequence of interest” (’012 patent) / “wherein the nucleic acid fragment comprising the nucleic acid sequence of interest” (’108 patent).....	159
E. Meyer Fails to Disclose the Other Limitations of the Independent Claims	163
F. It Would Not Have Been Obvious to Modify Meyer to Practice the Missing Claim Limitations	163
G. Meyer Fails to Anticipate or Render Obvious Claim 21 of the ’012 patent	164
H. Meyer Fails to Anticipate or Render Obvious Claim 22.....	167

<u>IX.</u>	Claim 2 is not obvious over Meyer in View of Siebert.....	172
X.	Claims 17-19 of the '012 patent and Claims 11-12, and 19 of the '108 patent are not obvious over Meyer in view of Bronner and/or Caruccio	178
XI.	SECONDARY CONSIDERATIONS.....	185
A.	Nexus.....	185
B.	Teach-Away/Skepticism	190
C.	Long-felt, unmet need	196
D.	Industry Praise for Tecan's Solution to these Long-Felt, Unmet Needs...	199
XII.	CONCLUSION	203

I, Carlos D. Bustamante, declare as follows:

1. I have been retained by Patent Owner, Tecan Genomics, Inc. (“Tecan”) as an expert witness in this *inter partes* review.

2. As part of my assignment, I have been asked to review Petitioner Integrated DNA Technologies, Inc. (“IDT”)’s petitions for *inter partes* review and to provide my opinions regarding positions taken by IDT.

3. I previously submitted a declaration as part of Patent Owner’s Preliminary Response. *See* EX2021. In that declaration, I considered IDT’s allegations that a person of ordinary skill in the art would not consider claim 1 of U.S. Patent No. 10,036,012 (“the ’012 patent,” EX1001) and claim 1 of U.S. Patent No. 10,876,108 (“the ’108 patent,” EX1002) to be novel over Meyer (EX1006). Because I determined that a POSA would consider the independent claims to be novel over Meyer, I also concluded that a POSA would find the challenged dependent claims of the ’012 and ’108 patents to be novel for at least the same reasons. I also considered IDT’s allegations that the challenged claims would have been obvious in light of Meyer. I found no reasons to support IDT’s claim that the claim limitations of the independent claims would have been obvious over Meyer. Because IDT failed to provide support for its claim that the independent claims are obvious, I also concluded that IDT failed to demonstrate obviousness of the dependent claims for the same reason. My opinions remain the same.

4. I understand that a panel from the United States Patent and Trademark Office's Patent Trial and Appeal Board ("PTAB") instituted *inter partes* review of the '012 and '108 patents. I have reviewed those institution decisions. This declaration expands on my prior testimony, addresses the issues raised in the institution decisions, and considers additional issues with IDT's claims. Where helpful, I have repeated relevant concepts discussed in EX2021.

5. In assessing novelty and obviousness, I have considered the contents of the '012 and '108 patents, the file histories of the patents and related patents, prior art publications, the two petitions for *inter partes* review, the related institution decisions, the Declaration of Dr. Eli Meyer (EX2287)—the first author of Meyer—and the Declaration of Peter A. Sims, Ph.D. (EX1085)—a declaration that IDT relies on to support its claims. I have considered the prior art relied on by IDT: Meyer (EX1006), Siebert (EX1007), Caruccio (EX1008), Bronner (EX1009), and Kelly (EX1010). I have also relied on my own experiences as a person of ordinary skill in the art as of the effective filing date of the '012 and '108 patents (January 26, 2012).

6. I have considered IDT's allegations that the challenged claims are anticipated or rendered obvious by Meyer (Grounds 1 and 2). As explained further below, I disagree with IDT's assessment. Meyer does not anticipate or render obvious the challenged claims of the '012 and '108 patents. The amplification process of Meyer that IDT relies on for its arguments does not result in the

enrichment of a target sequence. Instead of selectively amplifying only fragments with the sequence of interest (and thus enriching for the sequence of interest), the identified amplification process of Meyer amplifies cDNA fragments with and without the supposed sequence of interest (IDT argues that it is either the cap primer or the broken chain T primer sequence). This is because the method disclosed by Meyer was a solution for ensuring adequate representation of all sequences in the sequencing data of the coral transcriptome, not for the selective enrichment of specific, targeted sequences. Because Meyer fails to disclose an amplification scheme that enriches for a sequence of interest over other sequences in a sequencing library, I have concluded that Meyer fails to anticipate the challenged claims. I have concluded that Meyer further fails to anticipate claim 21 because Meyer does not disclose the fragmentation of RNA before reverse transcription. I have concluded that Meyer further fails to anticipate claim 22 because Meyer does not disclose ligating partial duplex adaptors to each end of the nucleic acid fragment containing the sequence of interest.

7. With respect to IDT's claims that Meyer renders obvious the claimed inventions, I have found no reasons to support IDT's claim that the claim limitations of the independent claims would have been obvious over Meyer. Because IDT fails to provide support for its claim that the independent claims are obvious, I have also

concluded that IDT has failed to demonstrate the obviousness of the dependent claims for the same reason, including for the claims challenged in Grounds 3-5.

8. I have also considered IDT's allegations that claim 2 would have been obvious over Meyer and Siebert (Ground 3). I disagree with IDT's claims that a POSA would have been motivated to adapt a method that relies on modifications to a reverse transcription process and apply that method for use with genomic DNA. I also disagree with IDT's claims that a POSA would have had a reasonable expectation of success in modifying Meyer for the claimed target enrichment of genomic DNA.

9. I have also considered IDT's allegations that claims 17-19 would have been obvious over Meyer and Bronner/Caruccio (Ground 5). I disagree with IDT's claims that a POSA would have been motivated to adapt a method specific to resolving issues with the 454 sequencer and apply that method to other NGS platforms. I also disagree with IDT's claims that a POSA would have had a reasonable expectation of success in modifying Meyer for the claimed target enrichment and sequencing with Illumina and Ion Torrent platforms.

10. I have also been asked to consider whether any objective indicia of nonobviousness apply to the claimed inventions. I have concluded that at least several objective indicia are applicable and that all support the nonobviousness of

the claimed inventions, which are embodied in Tecan's SPET products, including Tecan's Allegro® Targeted Genotyping products.

I. BACKGROUND AND QUALIFICATIONS

11. In this section, I summarize my educational background, career history, publications, and other relevant qualifications.

12. I have extensive experience in the interface of computational and biological sciences, as well as its application in medicine, agriculture, and evolutionary biology, spanning a career of over 25 years in the field. My education, work experiences, and research have focused on numerous relevant topics, including genomic sequencing and analysis, sample and library preparation for genomic sequencing, biostatistics, and genotype assay design. My expertise in these areas inform my opinions herein.

13. In 1997, I graduated magna cum laude, highest honors, with a Bachelor of Arts in Biology from Harvard University. Then, in 2001, I earned both my M.S. in Statistics and Ph.D. in Biology from Harvard University. From 2001 to 2002, I served as a postdoctoral researcher in mathematical genetics at the University of Oxford funded by a Marshall-Sherfield Scholarship.

14. I am currently an adjunct professor of Biomedical Data Science, Genetics, and (by courtesy) Biology at Stanford University where I have served

since 2010. I also served as the Inaugural Chair of the Department of Biomedical Data Sciences from 2015-2019.

15. I am a named inventor on numerous relevant patents and patent publications, including: U.S. Publication No. 2015/0360194 (“*Enrichment of DNA Sequencing Libraries from Samples Containing Small Amounts of Target DNA*”); U.S. Publication No. 2015/0232834 (“*Partitioning of DNA Sequencing Libraries into Host and Microbial Components*”); and U.S. Publication No. 2014/0045705 (“*Techniques for Determining Haplotype by Population Genotype and Sequence Data*”)—all of which are now issued patents. I have also completed numerous projects with the NIH, such as: *Inferring Multiple-SNP Disease Association with DNA Resequencing Data* (NIH/NIMH, 2008-2012); *Population Genetic Inferences from Dense Genotype Data* (NIH/NHGRI, 2010-2013); and *Methods for High-Resolution Analysis of Genetic Effects on Gene Expression* (NIH, 2013-2016).

16. In 2002, I joined Cornell University as an Assistant Professor in the Department of Biological Statistics and Computational Biology and Department of Statistical Sciences. I later became an Associate and Full tenured Professor at Cornell University, and also served as the Associate Director of the Cornell Center for Comparative and Population Genomics. At Cornell, my work primarily focused on population genetics and agricultural genomics.

17. I am the founder of Galatea Bio, Inc. and serve as President, CEO, and Board Member of the company. I served as a venture partner at Fidelity's early stage biotechnology group, F-Prime Capital, where the company was incubated during my tenure, and retain a venture partner title while serving as an operator of Galatea Bio. I am also the Principal and Chairman of CDB Consulting Group, Inc., an internationally recognized consulting firm, providing data science and genomics expertise in the areas of medicine, agriculture, and biology. I serve as a Board observer at Embark Veterinary, a company I advised at CDB Consulting and where F-Prime Capital is an investor.

18. I am also the founder of Arc-Bio (formerly IdentifyGenomics and BigData Bio) and now part of Cantata Bio. I served as chair of its Scientific Advisory Board and board member until the company was merged into Cantata Bio. Further, I was on the Board of Directors of Etalon Diagnostics and Eden Roc Sciences, two companies I helped start.

19. I have served on the advisory boards of various other corporations, non-profit entities, and government organizations. For example, I am currently a member of the scientific advisory boards at the following companies: Fauna Bio and Cantata Bio. I was also a member of the National Advisory Council to the National Human Genome Research Institute at the National Institutes of Health and remain on the Scientific Advisory Board of NHGRI Center for Research on Genomics and Global

Health. I was also the Co-Chair of the PAGE Analysis Working Group, the ClinGen Ancestry and Diversity Working Group, and the ClinGen Complex Disease/Polygenic Risk Score working group.

20. I have served on the Editorial Boards of numerous publications in the field, including PLoS-Genetics, American Journal of Human Genetics, Human Biology, Genome Research, Molecular Biology and Evolution, and Bioinformatics.

21. I have authored or co-authored over 250 articles and publications, including: *Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes*, 337 SCIENCE 6090 (2012); *Native American gene flow into Polynesia predating Easter Island settlement.*, 583(7817) NATURE 572-577 (2020); *Deconvoluting complex correlates of COVID-19 severity with a multi-omic pandemic tracking strategy*, NAT COMMUN (2022); *Polygenic risk score portability for common diseases across genetically diverse populations*, 18(1) HUM GENOMICS 93 (2024); *Neural ADMIXTURE for rapid genomic clustering*, 7 NAT COMPUT SCI. 621-629 (2023); *Validating and automating learning of cardiometabolic polygenic risk scores from direct-to-consumer genetic and phenotypic data: implications for scaling precision health research*, 16(1) HUM GENOMICS 37 (2022); *Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations*, 107(4) AM J HUM GENET. 788-789 (2020); *Fast and Cost-Effective Genetic Mapping in Apple Using Next-Generation Sequencing*, 4 G3 1681 (2014); *Achieving High-*

Sensitivity for Clinical Applications Using Augmented Exome Sequencing, 7 GENOME MED. 71 (2015); *A Research Roadmap for Next-Generation Sequencing Informatics*, 8 SCI. TRANSL. MED. 335 (2016); *Data Mining of Digitized Health Records in a Resource-Constrained Setting Reveals that Timely Immunophenotyping is Associated with Improved Breast Cancer Outcomes*, 18(1) BMC CANCER 933 (2018); *Genetic Analyses of Diverse Populations Improves Discovery for Complex Traits*, 570 NATURE 514 (2019); *Structural Variation Detection by Proximity Ligation from Formalin-Fixed, Paraffin-Embedded Tumor Tissue*, 21(3) J. MOLECULAR DIAGNOSTICS 375 (2019), *Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males versus Females*, 341 SCIENCE 562 (2013); and *Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences*, 48 NAT GENET. 593 (2016).

22. For my work in the field, I have received several honors and fellowships. These honors and fellowships include: the MacArthur “Genius” Award (2010), a Marshall-Sherfield Scholarship to Oxford (2001-2002), the Stanford Prize in Population Genetics and Society (2016); the Cornell Provost Award for Distinguished Research (2008); a Sloan Research Fellowship in Molecular Biology (2007-2009); and a Howard Hughes Medical Institute Graduate Research Fellowship (1999-2001).

23. A copy of my curriculum vitae, attached as **Exhibit A**, contains further details on my education, experience, publications, and the other qualifications allowing me to render an expert opinion in this matter.

24. I am being compensated my standard rate of \$800 per hour, plus expenses. My compensation is in no way dependent on the content of my testimony or upon the outcome of this proceeding.

25. In the past five years, I have testified as an expert witness in five other matters: (1) *Foundation Medicine Inc v. Guardant Health, Inc.*, IPR2019-00634 (P.T.A.B.); (2) *TwinStrand Biosciences, Inc. v. Guardant Health, Inc.*, C. A. 21-1126-GBW-SRF (D. Del.); (3) *State v Pease*, 88-10 (D. Mont.); (4) *Integrated DNA Technologies, Inc. v. Tecan Genomics, Inc.*, IPR2024-01504 (P.T.A.B.); and (5) *QIAGEN Sciences, LLC v. Tecan Genomics, Inc.*, IPR2025-00028, IPR2025-00029 (P.T.A.B.).

II. MATERIALS CONSIDERED

26. In formulating the opinions in this declaration, I have relied on over 25 years of work experience and knowledge regarding bioinformatics, genomics, and

nucleic acid sequencing. I also considered the documents cited in this declaration, including the Declaration of Dr. Eli Meyer.

III. LEGAL STANDARDS

27. I have been informed by counsel for Tecan of certain legal standards, and I have applied these standards in rendering my opinions.

28. When I say in this Declaration “I understand,” I mean that “I understand from counsel for Tecan in this IPR.”

29. I understand that to prove anticipation or obviousness in an *inter partes* review, it is the IDT’s ultimate burden to prove anticipation or obviousness by a preponderance of the evidence.

30. I understand that the question of whether the claims of a patent are anticipated or obvious is considered from the perspective of a person of ordinary skill in the art (the “POSA”).

31. I understand that the POSA is a legal construct representing a hypothetical individual with the same level of skill as the typical practitioner of the art at issue. I understand that courts consider a non-exhaustive list of factors, including (1) the educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) educational level

of active workers in the field. I understand that it is not appropriate to consider the level of skill possessed by the inventor in defining the POSA.

32. I understand that whether a patent claim is anticipated or obvious is ascertained at the relevant time period, which in this proceeding I understand to mean before January 26, 2012.

33. I understand that a prior art reference anticipates a patent claim if the prior art reference discloses subject matter within the scope of the claim, for all of the claim limitations.

34. I understand that an obviousness analysis involves ascertaining the scope and content of the prior art, the background and experience of a POSA, the differences between the claimed invention and the prior art, and whether there are additional factors present that may nevertheless weigh against obviousness such as unexpected results attributable to the invention actually claimed. I also understand that the obviousness analysis is based on the knowledge possessed by one of ordinary skill in the art at the time when the invention was made and that the use of hindsight is impermissible.

35. I understand that for a prior art reference, or a combination of references, to render a claimed invention obvious, the POSA must have had a motivation to modify or combine the references to arrive at the claimed invention. I also understand that separate from having a motivation to modify or combine, the

POSA must have also had a reasonable expectation of success in modifying or combining the references to arrive at the claimed invention.

36. I understand that the following additional factors are probative—and at times the most probative evidence—in assessing obviousness: (1) the prior art taught away from the claimed methods, (2) the art expressed skepticism towards the claimed inventions, (3) a long-felt, unmet need existed that was resolved by the claimed inventions, and (4) the industry praised the claimed inventions. I understand that these factors are called objective indicia of non-obviousness or secondary considerations. I understand that a prior art reference teaches away from the claimed invention when a POSA, upon reading the reference, would be led in a direction that diverges from the one taken by the applicant. I understand that the skepticism factor considers whether industry or a POSA expressed skepticism towards whether or how a problem could be solved. I also understand that skepticism may be shown where industry or a POSA express skepticism that the claimed invention would work. I understand that evidence demonstrating a long felt but unresolved need supports the non-obviousness of a claimed invention because it is reasonable to infer that the problem would not have persisted if it were obvious to address. I understand that industry praise of the claimed invention reflects that a POSA would not have reasonably expected the claimed invention to success.

37. I understand that there must be a nexus or connection between the claimed invention and the objective indicia of non-obviousness. I understand that nexus is presumed where the objective evidence is tied to a specific product that embodies the claimed features and is coextensive with them. I also understand that nexus may be shown where the objective evidence is the direct result of the unique characteristics of the claimed invention.

38. I understand that for a dependent claim to be anticipated or obvious, the independent claim from which that claim properly depends must be anticipated or obvious as well.

IV. TECHNICAL BACKGROUND

A. Genetic Information

39. All living organisms possess a genetic code, a chemical information system that contains the biological instructions required for an organism's development, function, and reproduction. Because an organism's genetic code dictates its biological functions, such as which proteins are produced, when they are produced, and where they are produced, understanding the composition and regulation of an organism's genetic information is essential to understanding biological function generally.

40. An organism's genetic information is contained in deoxyribonucleic acids ("DNA") and ribonucleic acids ("RNA"). These nucleic acids are polymers

of nucleotides—each nucleotide consists of a sugar, a phosphate, and one of four nitrogenous bases (A, C, G, or T/U)—and the order of these bases encodes biological information. For example, canonical DNA is composed of four bases: adenine (A), cytosine (C), guanine (G) and thymine (T). By organizing these bases into discrete linear sequences, nucleic acids can store information wherein the composition and order of the bases reflect a set of biological instructions that can be read by other cellular machinery.

41. Discrete sets of encoded information (such as the structure of a given protein) are retained in discrete regions termed “genes,” which are contained within a cell’s DNA. The totality of a cell’s DNA is referred to as a cell’s “genome.” DNA is organized in a helical structure composed of two intertwined strands. At a given location in the genome, each DNA strand is a “complement” of the opposite DNA strand to which it is associated, which allows for both strands of DNA to contain the same genetic information—the first sequence information contained in the “template” strand and the same sequence information mirrored in the complement strand. As I describe in further detail in subsequent sections, the complementary nature of nucleotides such as RNA and DNA is an integral feature exploited by sequencing technologies.

42. DNA is a molecule that encodes an organism’s genetic blueprint. DNA contains all of the information required to build and maintain an organism. Nearly

all nucleated cells within a single organism include exactly the same DNA. When required for actionable use within the cell, the genetic information contained in DNA is converted into RNA and then disseminated throughout the cell.

43. Like DNA, RNA is composed of nucleotides containing discrete bases: three bases—adenine (A), cytosine (C), and guanine (G)—shared with DNA and an additional base, uracil (U), corresponding to the thymine (T) base observed in DNA.

44. The genetic information contained in genes is converted to messenger RNA (mRNA) by an enzyme known as RNA polymerase during a process called transcription. mRNA is designed to efficiently traffic information from the DNA to sub-regions of the cell for processing into proteins. The entirety of a cell or organism's mRNA, including the genetic information contained therein, is referred to as its "transcriptome."

45. The structure of mRNA is critical for its function (such as during transcription, processing, translation, and eventual degradation). Notably, the mRNA structure includes:

- a. ***The "5' Cap."*** At the 5' end of the mRNA, there is a modified nucleotide cap, known as the 5' cap. This cap is linked to the mRNA via a unique 5' to 5' triphosphate linkage. The 5' cap is crucial for the stability of the mRNA and plays an important role in the initiation of translation. It also protects the mRNA from degradation by exonucleases.

b. ***The coding sequence of the mRNA.*** The central region of the mRNA molecule contains the coding sequence which carries the genetic information encoded in the form of codons. Each codon specifies a particular amino acid. This sequence is read by the ribosome in the 5' to 3' direction during the process known as translation. The ribosome assembles a protein based on the mRNA coding sequence.

c. ***The polyA tail.*** The polyA tail is a long chain of adenine nucleotides that is added to an mRNA molecule during RNA processing to increase the stability of the molecule. Immediately after a gene in a eukaryotic cell is transcribed, the new RNA molecule undergoes several modifications known as RNA processing. These modifications alter both ends of the primary RNA transcript to produce a mature mRNA molecule. The processing of the 3' end adds a polyA tail to the RNA molecule. The polyA tail can be between 100 and 250 bases long. The polyA tail makes the RNA molecule more stable and prevents its degradation. Additionally, the polyA tail allows the mature mRNA molecule to be exported from the nucleus and translated into a protein by ribosomes in the cytoplasm.

46. Complementary DNA (cDNA) is a form of DNA that is synthesized from an mRNA template in a process called reverse transcription. Because the RNA template is typically mature, spliced mRNA, the resulting cDNA represents mainly

exon sequences (those sequences that code for protein) and therefore mirrors the set of genes actively expressed in the sampled cells or tissue. This makes cDNA especially useful in genetic studies because it typically represents the expressed genes in a cell, tissue, or organism, excluding introns and regulatory sequences found in genomic DNA.

47. In addition to studying the typical sequences of nucleic acids, researchers are also interested in atypical sequences. Accidental alterations in an organism's genetic sequence, termed "mutations," are associated with and often drive numerous diseases. For example, single base changes in an organism's DNA, known as "single nucleotide polymorphisms" (SNPs), can result in major downstream changes in protein expression and function, and in some cases can be the main driver of disease, such as sickle cell anemia. In additional cases known as "complex diseases," multiple SNPs or similar mutations throughout the genome combine to drive or exacerbate a disease state.

B. Nucleic Acid Genotyping

48. Genetic variation within a population arises when changes in an organism's DNA sequence occur that are then passed down to their descendants. When a specific variant is detected at a considerable frequency in a population—typically above 1%—it is termed a "genetic polymorphism." In contrast, "mutations" usually refer to the initial rare or novel alterations that arise

spontaneously and can be associated with genetic diseases, but may or may not be passed down to descendants or may arise in a specialized setting (such as a cancerous tumor). A SNP represents one of the most prevalent types of polymorphisms. Though many SNPs are clinically benign, some can markedly affect gene expression and protein function, potentially driving or contributing to disease, as exemplified by the single-base mutation in the beta-globin gene (HBB) responsible for sickle cell anemia, which has been known since the 1950s to be a disease with a genetic origin and etiology.

49. To systematically identify and characterize these variations, researchers developed the process of “nucleic acid genotyping,” which involves examining a segment of DNA and comparing it to a reference sequence or to the same segment in another individual. In diploid organisms—those that inherit two sets of chromosomes, one from each parent—each locus is typically present in two copies.¹ Therefore, an individual’s diploid genotype at a specific SNP site can be homozygous reference [AA], heterozygous [AB], or homozygous non-reference [BB], where “A” denotes the reference (or “wild type”) allele and “B” denotes the

¹ The exceptions are uniparental inheritance of the mitochondria and Y chromosome, which are only passed on from the maternal or paternal parent and are considered haploid.

non-reference (or “minor”) allele. By systematically determining whether an individual has [AA], [AB], or [BB] at many such loci, scientists can create a detailed map of genetic variation relevant to traits or diseases.

50. This locus-specific approach enabled researchers to organize and interpret genetic data on a large scale, thereby identifying regions of the genome that play pivotal roles in health and disease. Because SNPs and other polymorphisms can provide insight into disease susceptibility, severity, and treatment response, scientists and clinicians benefit from methods that can genotype many loci simultaneously. To meet this need, high-throughput platforms were developed, including Affymetrix’s large-scale genotyping arrays, which rapidly evaluate tens of thousands—or even millions—of SNP positions in a single experiment. By focusing on known sites of variation across the genome, these arrays generate substantial volumes of data that can be used to identify disease-associated alleles (via genome-wide association studies), pinpoint genetic markers for clinical diagnostics, and facilitate personalized medicine.

51. Some of the earliest efforts to characterize genomic information included genotyping through genome walking, DNA microarrays, and sequencing.

C. Library Preparation

52. Before I discuss how prior art genome walking, DNA microarrays, and sequencing technologies worked, it is important to consider how genetic samples were prepared for use with these techniques.

53. Generally, after sample collection, researchers isolate the sample to be analyzed and prepare a library of that sample for further analysis. The procedures used to generate that library are referred to as “library preparation.” Library preparation methods can vary for a number of reasons, but there is a general workflow when analyzing samples through genome walking, DNA hybridization arrays, and sequencers. Historically, this general workflow has consisted of fragmentation, adapter ligation, and amplification.

54. Fragmentation refers to a specific procedure in library preparation where larger samples are broken into smaller fragments through either physical or enzymatic means. Examples of physical fragmentation include nebulization (using compressed nitrogen or air to force samples through small holes), sonication (breaking covalent bonds with ultrasonic waves), and hydrodynamic shearing (using an aqueous solution to force samples through constrictions). *See* EX1001.0018 (6:60-63). There are numerous reasons why a researcher would fragment the sample during library preparation, including that the instruments used for genetic analysis typically only work with nucleic acid fragments. For example, Illumina sequence

reads are governed by cycle lengths of typically 50, 75, or 150 nucleotides, yielding paired reads of (2x50, 2x75, or 2x150 base pairs in length). Typical fragment length was therefore 300-500 base pairs in length since at most the first (and last) 50, 75, or 150 base pairs of the fragment are sequenced by the instrument. Another reason is that the fragmentation process can separate fragments of interest from other genetic material, such as with enzymatic fragmentation, which targets specific restriction sites (e.g., in the use for genotype-by-sequencing assessing genetic polymorphism).

55. Adapter ligation refers to attachment of adapters to the nucleic acid sample or fragments. There are numerous purposes for adapters, including for extension, amplification, and annealing fragments to instruments for analysis (such as the Illumina P5/P7 sequences which permit flow-cell binding).

56. Amplification refers to the increase of samples or fragments. It can be linear or exponential. It can also be conducted through thermal cycling or isothermal.

57. One method for amplification is polymerase chain reaction or PCR. Conceived by Kary Mullis in 1983, PCR utilizes enzymes known as polymerases and short oligonucleotides known as primers to catalyze the replication of new nucleic acid strands from a provided nucleic acid template. EX2040.0001. PCR accomplishes this replication through a series of reactions where the complement of

a nucleic acid strand is generated. *Id.* PCR proceeds through iterative temperature cycling—denaturation, annealing, and extension—that conceptually doubles the number of target molecules with each cycle. *Id.*, 0003-05. This cyclical approach allows for exponential amplification of the nucleic acid strand. *Id.*, .0001.

58. In PCR, two primers are designed to flank the nucleic acid template, with the nucleotide sequences of the primers determining the reaction’s specificity. *Id.*, .0003-05. The first primer binds to a target site either located on the nucleic acid template or as part of an appended adapter, a short synthetic oligonucleotide that can be attached to the nucleic acid template. *Id.*; EX1009.0004. This process is known as annealing. The polymerase then extends from the 3' end of the primer by sequentially incorporating deoxyribonucleotide triphosphates (“dNTPs,” a type of nucleotide) that are complementary to a provided template strand. EX2040.0001. This process is known as extension, and generates a nucleotide fragment that contains the primer and is complementary to the strand from which it is synthesized. *Id.*, .0003-05. The extension product (as well as the complementary sequence of the original fragment) can then be bound by the second primer, and the process repeats itself, generating a second nucleotide fragment identical to the original template strand. *Id.* Over many PCR cycles, often 25–35 cycles, an initial nucleotide fragment can be exponentially amplified. *Id.*, .0001. The resulting products will

contain the original nucleotide fragments along with an exponential number of copies of each fragment. These copies are often referred to as “amplicons.”

59. PCR provided researchers with the ability to amplify low amounts of genetic material, providing access to genetic material that occurs in low amounts and replenishing genetic material that could then be used for further experiments and methods. EX1009.0004. As primer sequences are introduced into the nucleotide fragments as part of the PCR process, specific synthetic sequences can be incorporated into the primers such that the synthetic sequences are introduced into the population of PCR products. *Id.* This allows researchers to manipulate PCR libraries to prepare them for downstream processing, such as introducing primer binding sites for sequencing primers. *Id.*, .0004-05.

60. While PCR expanded the capabilities of the field, it was also considered to be error prone. EX2075.0004; EX2041.0005; EX2036.0009-10 (31:20-33:13). Any one of several factors—including primers, polymerases, reaction temperature, and amplification cycles—contributed to the generation of off-products and other artifacts. EX2042.0008-11. For example, despite the development of high-fidelity polymerases exhibiting lower error rates, hundreds of millions of extension events across 30 cycles inevitably introduce accidental nucleotide substitutions and single-nucleotide insertions and deletions, into PCR products. EX2075.0004. Additionally, phenomena such as sequence-content bias and polymerase slippage and premature

dissociation can lead to under- or over-representation of select PCR products or erroneous products. EX2042.0001. These errors are exponentially multiplied as more cycles of PCR occur, resulting in populations of amplified products containing many artificial sequence errors introduced as part of the PCR process. Many of these erroneous products are referred to as “artifacts.”

61. Researchers attempted to minimize the errors associated with PCR. For example, numerous methods of PCR were developed as part of “genome walking,” including methods that introduced multiple sets of PCR cycles, additional components (like biotin) that could be used to capture fragments, and specialized adapters and primers that could suppress PCR amplification. *See, e.g.,* EX1007.0001. In general, each method demonstrated an improvement on the prior methods, but as explained in the following section, even later generation of methods still suffered from target specificity issues and amplification artifacts.

D. Anchored PCR and Genome Walking

62. The invention of PCR gave researchers the ability to selectively target genetic regions and amplify them. EX2040.0001. As early as the 1990s, researchers coopted this facet of PCR to develop methods to explore unknown regions of the genome, a technique known as “anchored PCR” or “genome walking” (when applied to genomic DNA). EX1007.0001. Researchers designed primers to target two known nucleotide sequences that flanked an unknown genetic sequence. *Id.*

Typically, one of the known sequences was either previously characterized endogenous sequences (sequences naturally present in the genetic sample of interest) while the other was a synthetic nucleotide sequences introduced into the sample, for example by an adapter that was attached to a nucleotide fragment containing the unknown genetic sequence. *Id.*, EX2044.0001. Using PCR, the unknown genetic sequence in between the two primer sites could be amplified for further characterization. *Id.* Similar methods existed for RNA, which utilized a technique known as rapid amplification of cDNA ends (“RACE,” which converts RNA to cDNA) to introduce known sequences at the end of the resulting molecule. EX2043.0001. These known sequences could then be targeted in conjunction with a second primer targeting an internal portion of the cDNA. *EXId.*, .0003.

63. Due to their design, anchored PCR and genome walking suffered from serious issues that hampered their effectiveness. As PCR-based methods that used primers to target specific regions of interest in a larger genetic sample, anchored PCR and genome walking relied on the specificity of the primers to accurately amplify the region of interest. EX1007.0001; EX2047.0001. Primer specificity refers to the ability of a given primer to bind to its intended target nucleotide sequence, as opposed to a) binding non-specifically to an unintended nucleotide sequence or b) binding the targeted nucleotide sequence at multiple locations in the sample, as nucleotide sequences may be repeated in genetic samples at different

locations. The importance of primer specificity is compounded when using a synthetic sequence, such as an adapter as a primer binding sequence, as many methods introduce the synthetic sequence non-discriminately into the sample. *Id.* In these cases, only the primer targeting the second, endogenous nucleotide sequence can be expected to confer specificity. EX2047.0001.

64. Non-specific primer binding, referred to as “off-target binding” is a major problem in genome walking because it leads to the amplification of unwanted DNA fragments, obscuring the desired target sequence.² *Id.* This happens because primers can anneal to multiple locations in the genome, not just the intended target site, and with exponential amplification in PCR, these non-specific products can quickly overwhelm the specific product. *Id.* This makes it difficult to isolate and identify the specific DNA sequence of interest. Early anchored PCR and genome walking methods suffered from serious issues with off-target binding. EX2043.0002; EX2047.0001. Off-target binding was particularly problematic when primers targeted repetitive sequence elements, which are portions of nucleotide sequences that occur repeatedly throughout a sample. EX2043.0010. Additionally, primer binding and PCR generally suffered from issues with certain nucleotide-rich

² This phenomenon, when discussed in the context of PCR amplification, is also known as “non-specific” amplification.

sequences (e.g., A-T or G-C rich), which further narrowed the applicability of target-specific PCR methods such as anchored PCR and genome walking. *Id.*

65. To address the issue of primer specificity, researchers modified the methods to optimize the PCR process to and limit the generation of off-target products. EX2043.0002; EX2045.0001. By the late 1990s, several methods existed that were designed to minimize non-specific amplification, including (1) PCR suppression, where non-desired PCR products are “suppressed” and not subsequently amplified further; (2) vectorette PCR, which utilizes special adapters with mismatched sequences that blocks amplification where a fill-in reaction does not correct for the mismatch; (3) splinkerette PCR, which is similar to vectorette PCR but uses a hairpin loop structure on the adapter in an attempt to promote further on-target specificity; and (4) nested PCR, a method wherein a second round of PCR is conducted using primers designed to bind selectively to the product of the first round of PCR. EX2044.0001; EX2045.0001; EX2046.0001. Generally, these methods were designed to either a) limit the ability of off-target products generated in a first round of PCR to be further amplified or b) add further rounds of primer targeting using additional primers to improve the specificity of the resulting PCR product. *Id.* Additional methods, such as adaptor long-range PCR, were also developed to further refine primer specificity-driven PCR in an effort to limit off-target products. EX2045.0001.

66. Most of these early methods relied on a combination of a first primer which targets a universal synthetic primer binding site introduced into the sample, a second primer which targets a portion of the sample itself, and some additional feature to control for amplification. EX2043.0003; EX2044.0001; EX2045.0001; EX2046.0001. While these early, pre-NGS methods may have used a single target-specific primer, they demonstrated why the field had moved beyond single target-specific primers by January 2012:

- In 1995, Siebert *et al.* observed that “unpredictably primed PCR” (a method of characterizing a genetic sample by randomly annealing primers targeting common repetitive sequence elements and conducting PCR) and vectorette PCR “generated multiple PCR products even after nested PCR was performed.” EX1007.0001. Siebert explained that “the presence of multiple PCR products would complicate further characterization without Southern blot hybridization or extensive cloning.” *Id.* In an attempt to resolve this issue, Siebert developed a method that coupled vectorette PCR with PCR suppression. This method relied on special, vectorette primers, that were ligated to both sides of a DNA fragment. The vectorette primers contained a suppression tag that would bend fragments into a hairpin loop, suppressing exponential amplification. I discuss this further below. *See infra* ¶¶ 200-204. Siebert determined that this method improved target specificity but that the

specificity was “due to the suppression effect of the adaptor/adaptor primer combination.” EX1007.0002.

- In 1996, Chenchik *et al.* (including Siebert) disclosed a method to a RACE protocol designed to selectively target the ends of cDNA molecules. EX2043.0001. The method introduced adapters at the end of the cDNA molecule that could be used in conjunction with a primer targeting an internal sequence to PCR the ends of the fragments. *Id.* The adapters were designed to only function in conjunction with the target-specific primer, and would suppress any fragment that contained two full-length adapter sequences. *Id.*, .0003. The method was further combined with a long and accurate PCR method. *Id.*, .0001. Chenchik observed that prior RACE methods suffered from “high background of nonspecific PCR products.” *Id.* In a proof of concept experiment, Chenchik observed a reduction, but not elimination, in nonspecific PCR products. *Id.*, .0004-06.
- In 1998, Padegimas *et al.* reported an adapter ligation-based walking technique designed to improve PCR specificity using an optimized adapter ligation strategy while removing (via degradation) un-ligated fragments. EX2045.0001. The adapters were designed with three discrete structural components to enhance stability and prevent degradation during the newly implemented degradation step, wherein enzymes were added to degrade non-

ligated nucleic acid fragments in an effort to enhance PCR specificity. *Id.*, .0003, 05. Following adapter ligation and non-target degradation, primers would then be used to target the adapters and a target sequence contained within the ligated nucleic acid fragments for PCR. *Id.*, .0003. Padegimas states that the method was developed to address “low specificity, low efficiency, and/or extensive cycling that contributed to a higher error rate in PCR products” in current PCR based walking methods, including vectorette, splinkerette, and adapter-based methods. *Id.*, .0001. Padegimas reported a reduction in off-target products with its method, although the method required high degree of optimization to achieve this reduction. *Id.*, .0004.

- In 2005, Miura *et al.* reported that “PCR applications [that] use only a single unique primer in combination with a common primer [] includ[ing] rapid amplification of cDNA ends (RACE) [], PCR-mediated genome walking [], adaptor-tagged competitive PCR (ATAC-PCR) [], and so forth” were “often plagued with non-specific amplification products derived from so-called ‘off-targets.’” EX2047.0001. To address this, Miura sought to develop an algorithm to determine primer binding specificity performance. *Id.* In two proof-of-concept experiments, Miura observed that off-target priming could be reduced when primer design was optimized, but not eliminated. *Id.*, .0004-

06. Further, Miura observed that off-target primer binding was sometimes exacerbated using the algorithmic design. *Id.*, .0007.

- Also in 2005, Tsoktouridis *et al.* compared various PCR methods and found “no simplicity, low reproducibility, high backgrounds of nonspecific PCR products, insufficient homopolymer tailing, small PCR product amplification, and laborious time-consuming procedures.” EX2048.0001. As Tsoktouridis explained, techniques “using adaptors for PCR amplification of unknown DNA fragments flanking a known sequence” (genome walking methods) were “limited by a relative short walking distance, low specificity, low efficiency, and/or extensive cycling, which contributed to a higher error rate in PCR products.” *Id.*, .0003. To circumvent these issues, Tsoktouridis developed an optimized adaptor long-range PCR method, wherein adapters were designed for use with specific digestion enzymes (enzymes designed to cut DNA at certain locations) after which long-range PCR could be conducted using a primer that targeted the adapters. *Id.*, .0003-04. Additionally, Tsoktouridis observed that further optimization of the amount of sample used with the methods was required to reduce off-target primer binding. *Id.*, .0004.
- In 2011, researchers recognized that despite that “many studies have been made to improve the genome walking method,” “it is still in need of improvement.” *Id.*, .0002-03. Sohn *et al.*, commenting on the field in 2011,

explained that the “most serious problems of the PCR-based genome walking methods are low amplification efficiency and non-specific amplification.” *Id.* Sohn noted that to “improve this problem, complicated multi-step methods were developed, including secondary PCR, purification of desired PCR products using biotinylated primers, or improvement of PCR cycle conditions” however, even then, “satisfactory performance [had] not yet been achieved.” *Id.* Sohn proposed yet another modification to the traditional genome walking method, utilizing a nucleotide analog ligation approach coupled with PCR, wherein the ends of the nucleic acid fragments in a sample were chemically modified to introduce a nucleotide chemical analog that could be target to enhance adapter ligation. *Id.*, .0002. after which the adapter could be used in conjunction with a target specific primer to target via PCR a region of interest within the fragment. *Id.*

67. Indeed, as demonstrated above, despite over sixteen years of method development, primer specificity remained a serious issue limiting the utility of primer-specific PCR methods. Over time, the art moved away from using single target-specific primers without many additional safeguards, such as complex PCR suppression schemes. EX2048.0002-03. By the mid 2000s, algorithmic design approaches were also being developed to optimize primer design in an effort to reduce primer off-target binding. EX2047.0001. However, even these efforts failed

to resolve off-target issues, and it was generally accepted that primer specificity issues had to be mitigated by increasingly complex PCR assays and optimized primer design. EX2048.0002-03. As such, off-target primer binding remained an unresolved issue inherent in PCR applications. Indeed, by the late 2000s, researchers were considering alternatives to PCR amplification to circumvent these issues and enable multiplex sequencing of targeted gene regions.

E. PCR-Free Sequencing Methods

68. As discussed above, despite the tremendous utility of PCR, issues with logistics, sequence errors, primer specificity, and amplification bias limited its application. EX2042.0001-02; EX2051.0001, 0003; EX2052.0001-02. These issues were particularly pronounced in applications assessing the composition of a genetic sequence, as these PCR issues could convolute or artificially bias the resulting sequence analysis. EX2042.0001-02; EX2052.0001-02. With the advent of next-generation sequencing methods (discussed in further detail below), PCR issues became even more pronounced, as many NGS methods utilized PCR for both sample preparation and as part of the sequencing process. *Id.* Further, as the scale of sequencing methods increased, inherent limitations in PCR applications became apparent as primer-specific approaches were not considered scalable to whole genome assessments. EX2051.0001, 0003; EX2052.0001-02; EX2066.0001.

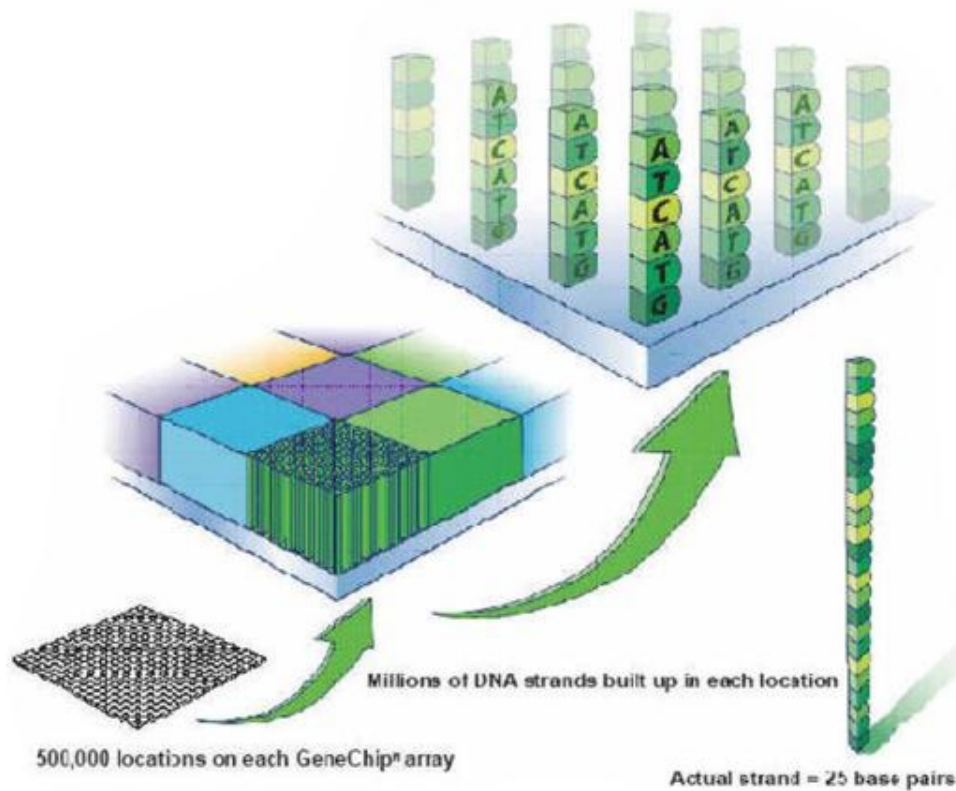
69. Given these issues, many researchers began to consider alternative approaches for handling genetic material. *Id.* Several proposed incorporating required sequencing components directly into adapters in order to eliminate the need to introduce them during PCR (allowing researchers to forego PCR altogether). EX1009.0019; EX2053.0003-04; EX2054.0001-02. Methods were also developed to minimize the amount of PCR amplification required prior to sequencing by replacing certain PCR steps with selective capture methods (molecular techniques designed to capture sequences of interest in a larger pool of sequences for downstream assessment). EX2052.0001-02. Still other researchers sought novel hybrid array methods, which were plates designed to capture sequences of interest for assessment, without any PCR amplification step. *Id.*, .0011. By 2011 PCR amplification was widely recognized as the predominant source of error in standard sequencing workflows. EX2042.0001-02. Indeed, issues stemming from with PCR amplification in NGS sequencing were such that companies developed commercial sample preparation kits designed to limit PCR amplification during sample preparation for NGS applications. EX1008.0001; EX2056.0001.

F. Hybridization Arrays

70. One key high-throughput genotyping approach involved hybridization arrays, which exploit the principle that complementary nucleic acid strands selectively bind to one another (e.g., adenine pairing with thymine, guanine pairing

with cytosine). In the early 2000s, Affymetrix developed a photolithographic manufacturing process—inspired by semiconductor fabrication—to synthesize short 25-mer oligonucleotide probes directly on a silica or glass substrate. Through iterative cycles of UV light exposure through custom-designed photomasks, nucleotides were selectively deprotected and added in a precise, stepwise manner, creating microscopic “features,” each containing up to a million copies of a single probe sequence. On arrays designed to genotype more than 10,000 SNPs, each SNP was interrogated by over 40 probes, incorporating perfect-match and mismatch designs. Consequently, a single array featured over 400,000 distinct probe sequences, which—multiplied by the millions of replicated copies per feature—resulted in hundreds of billions of physical probe molecules on one chip. During a genotyping assay, only DNA fragments complementary to these probes would hybridize and remain on the array after washing, enabling accurate, parallel determination of whether an individual was homozygous reference [AA], heterozygous [AB], or homozygous non-reference [BB] at each interrogated locus. Affymetrix thus enabled high-throughput, highly specific genotyping (~99.8% accuracy) across tens to hundreds of thousands of highly polymorphic SNPs.

71. A schematic depiction of the surface of an array is shown below:



EX2003.0001. Because each probe was synthesized at a known, fixed location on the array, researchers could interpret hybridization patterns using specialized software to identify genes, relative expression levels (for expression arrays), or genotypes at specific SNPs (for genotyping arrays).

72. While genotyping by hybridization arrays provided a powerful, high-throughput platform for genetic assessment, the arrays were inherently targeted in nature. Probes were designed with high specificity to known target sequences, so the arrays could only query the particular SNPs for which they had been manufactured. Additionally, in order to design SNP probes, the location and flanking

sequence of the SNP had to be known in advance. This requirement limited the applicability of microarrays in exploratory genomics where the relevant sequences or variants were unknown, driving the development of next-generation sequencing approaches that permit *de novo* discovery of genomic alterations.

G. Nucleic Acid Sequencing

1. Pre-NGS Technologies for Sequencing

73. Given the role of the genetic code in governing biological function, developing methods of identifying genetic sequences in various biological contexts has been a primary goal of scientific research. Early methods focused on DNA sequencing and technical methods for “reading out” the nucleotide composition of individual DNA strands. Early sequencing methods were time- and resource-consuming and were not suited to scalable applications. These limitations made sequencing large polynucleotides a technical challenge and genomes, basically infeasible. These issues were somewhat ameliorated by the advent of large scale-cloning and/or the parallel sequencing of many regions of DNA that would then be assembled together into a “working draft” of the genome.

74. One of the first sequencing methods, “Sanger Sequencing,” also known as the chain termination method, was developed by Frederick Sanger in 1977. Before the advent of Polymerase Chain Reaction (“PCR”) technology, Sanger Sequencing was performed directly on purified DNA templates, typically from

cloned DNA fragments. Sanger Sequencing involved the enzymatic synthesis of DNA strands and their controlled termination using dideoxynucleotides (“ddNTPs”).

75. There were major challenges and limitations to this approach. The need to clone and grow bacterial cultures to obtain sufficient DNA for sequencing made the process cumbersome and slow. The quality and length of the sequence reads were limited by the efficiency of the cloning process and the resolution of the electrophoresis gel. The method was also less sensitive and less scalable compared to later techniques that utilized PCR for amplification, making it challenging to sequence very small amounts of DNA or conduct large-scale sequencing projects efficiently.

76. The advent of PCR revolutionized Sanger Sequencing. Through PCR, researchers were able to take nucleic acid samples and amplify those samples without the need to clone and grow bacterial cultures. By running a series of PCR cycles, the original DNA sample can be exponentially amplified, as the strands produced in each PCR cycle can then serve as template strands in the following cycle. *See* EX2001.0001.

77. Another innovation to Sanger Sequencing was its automation through platforms such as the ABI 3730 DNA Analyzer, which transformed the landscape of genetic analysis by integrating advanced automation and fluorescence-based detection technologies. Sequencing samples were prepared in a mix that included

the purified DNA template, a sequencing primer, DNA polymerase, and a mixture of regular dNTPs alongside specially dye-labeled dideoxynucleotides (“ddNTPs”). Each type of ddNTP, which terminates DNA chain elongation upon incorporation, was tagged with a distinct fluorescent dye. This labeling allowed for the simultaneous identification of each nucleotide during the sequencing process, despite the reactions occurring in a single tube. The data collected from the ABI 3730 was processed using software that translated the signals into a readable sequence displayed as a chromatogram, which could be interpreted to identify the DNA sequence. This automated system reduced the manual effort involved in sequencing and increased both throughput and accuracy.

78. Sanger Sequencing was limited to sequencing a single DNA region of roughly 500–900 base-pairs per sequencing reaction. Even with the advent of automated Sanger Sequencing, this structural aspect of Sanger Sequencing limited its sequencing throughput, making sequencing at scale both costly and logistically challenging. For example, completion of the Human Genome Project took approximately 13 years and 2.7 B USD, while the National Human Genome Research Institute estimated that sequencing a second human genome using the Sanger Sequencing methods available in 2003 would cost another 50-100 M USD. *See* EX2002.0001, 06. These technical limitations rendered general application of Sanger Sequencing to large or complex genomic sequencing tasks extremely

expensive and time consuming, limiting genome-scale activity (in the US at least) to a dozen or so nationally-designated sequencing centers funded through large infrastructure grants from the National Institutes of Health, Department of Energy, and the like.

2. Next Generation Sequencing

79. Coinciding with the completion of the Human Genome Project in the early 2000s, great scientific and public interest in a deeper understanding of the genome led to the commercial development of more affordable and scalable techniques for genomic assessment. These efforts resulted in a series of novel technologies and methods collectively termed “Next-Generation Sequencing (‘NGS’) Technologies.” EX2003.0001. As compared to traditional sequencing methods such as Sanger Sequencing, NGS technologies were characterized as cheaper, more efficient, and capable of assessing much larger aggregate portions of the genome despite shorter individual read lengths. As compared to prior methods, NGS sequencing was defined in part by its “massively parallel” nature—*i.e.* its ability to process tens of thousands to millions of DNA sequences at the same time. Most commercial NGS technologies (e.g., 454, Illumina, Ion Torrent) utilized a “sequencing-by-synthesis (SBS)” strategy. EX2004.0001-02. SBS leveraged the sequential nature of genomic information to assess each nucleotide in a sequence in

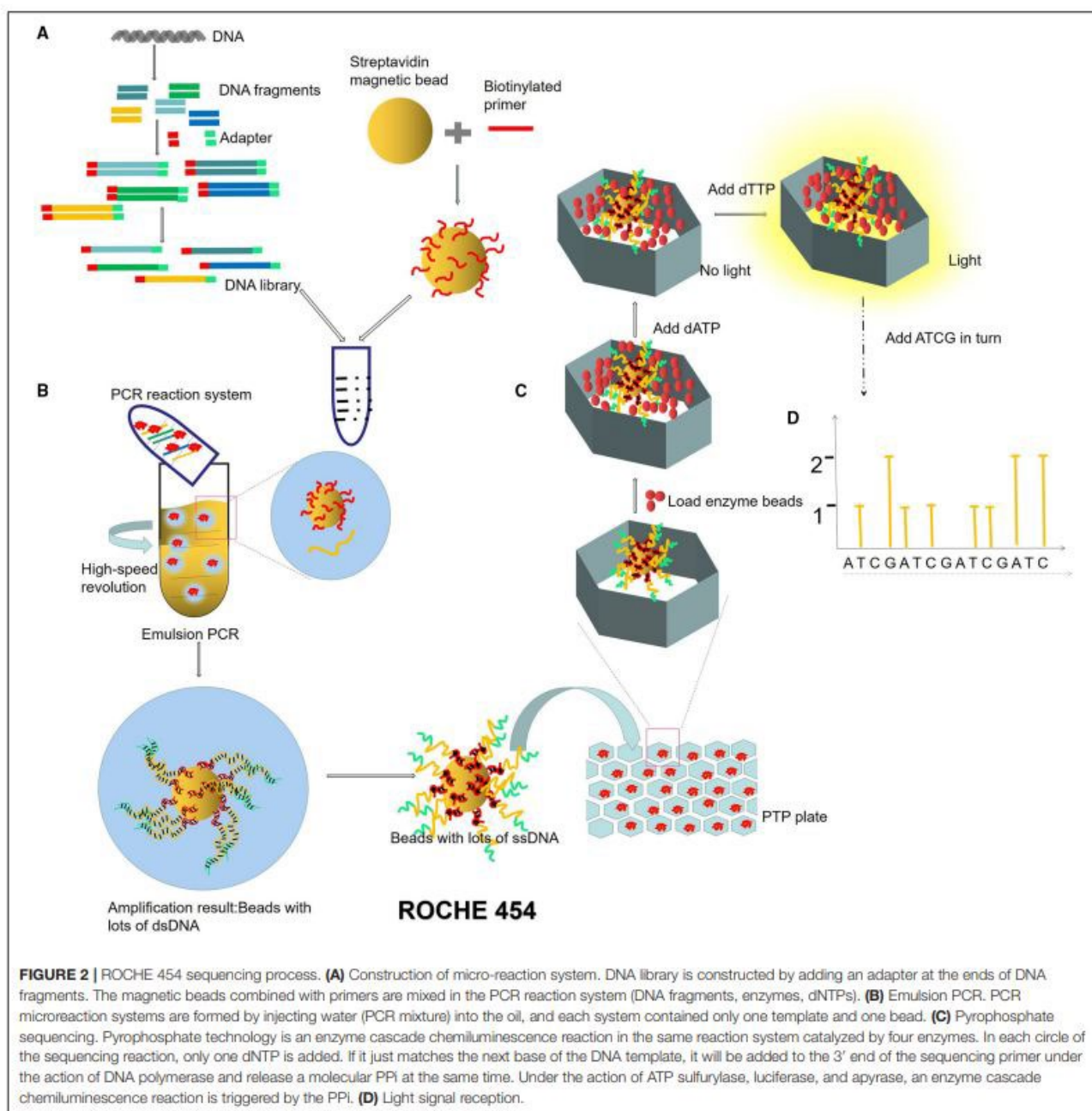
order. Polymerases were coupled to a signaling output that allowed for the identification of the incorporated nucleotide.

80. Compared to traditional sequencing methods, early NGS technology afforded researchers several benefits. Due to their design, NGS systems allowed researchers to sequence thousands of reads in tandem in an automated manner, increasing both the platform capacity and the depth and coverage of the generated sequencing data. This made genetic sequencing both cheaper and more scalable as compared to prior methods. The parallel sequencing paradigm allowed researchers to sequence multiple genes and assess millions of variants simultaneously, while reducing the required DNA input and, due to the scaled approach, decreasing sequencing times. These factors led to an overwhelming increase in the amount of sequencing data generated in the course of research and a deeper understanding of the structure and content of many genomes across species. EX2003.0001.

81. Pyrosequencers, such as the 454 sequencers, were among some of the first set of platforms that harnessed this “second generation” sequencing technology. Illumina produced the most commercially successful sequencing instruments in this class, which also included Ion Torrent, Helicos, and ABI/Lifetech SOLiD (among others). Over time, NGS sequencers improved on their ability to process longer read lengths, which resulted in a “third generation” of sequencers, which included Nanopore and PacBio sequencers.

82. Around January 2012, 454 sequencers, Illumina, and SOLiD were considered to be the three platforms that were “widely available through academic core laboratories and commercial service providers.” EX2057.0003. These three platforms “traditionally split their focus into fewer long reads (454) vs. more short reads (Illumina and SOLiD).” *Id.* A POSA would have understood that long reads were “optimal for initial genome and transcriptome characterization because longer pieces assemble more efficiently than shorter pieces” but that shorter reads would have been favored for more targeted sequencing, especially given the diminished accuracy of sequencing longer reads at the time. *Id.*

83. The sequencing results reported in Meyer were generated with the 454 GS-Flx sequencing platform. EX1006.0001. The 454 GS-Flx was released in 2007 and represented an improvement over the GS20 sequencing platform, which had been released in 2005. The 454 GS-Flx could sequence millions of nucleic acid fragments in a single run. As can be seen from the figure below, there were generally four stages to sequencing with the 454 platform.



EX2058.0005. While this figure was not published until years after the discontinuation of the 454 sequencing platforms, it is an accurate representation of the same four-step workflow depicted in the October 2008 GS-Flx Operating Manual. See EX2059.0020.

84. In the first stage (A), the micro-reaction system was constructed. This stage involved two distinct processes. First, library preparation occurred. Library preparation varied but generally involved the fragmentation of genetic material and the ligation of universal 454 adapters to the fragments. Second, the prepared library was combined with biotinylated streptavidin magnetic beads. The biotin captured one fragment per bead.

85. In the second stage (B), emulsion PCR (“emPCR”) occurred. The streptavidin beads and captured fragments of the last step were added to a reaction mixture where each bead was separately immersed in oil. Each emulsion functioned as its own PCR reaction. The fragment on the emulsified bead was amplified to exponentially increase the copies of that fragment.

86. In the third stage (C), pyrophosphate sequencing (i.e. pyrosequencing) occurred. The amplified beads were rolled across a picotiter plate that contained over a million wells. A single bead would fall into each well. From there, complementary strands of the captured fragments would be generated. For this process, a series of washes would occur where a single species of nucleotide (adenine, thymine, guanine, or cytosine) would be washed across the plate. If that nucleotide was the next nucleotide in the complementary sequence, then it would bind to the nascent strand. Specific enzymes in the wells would facilitate a chemiluminescence reaction during strand elongation, emitting light as a result of

the binding. This process would be repeated as numerous rounds of washing occurred for the different nucleotide species. The figure below illustrates this process.

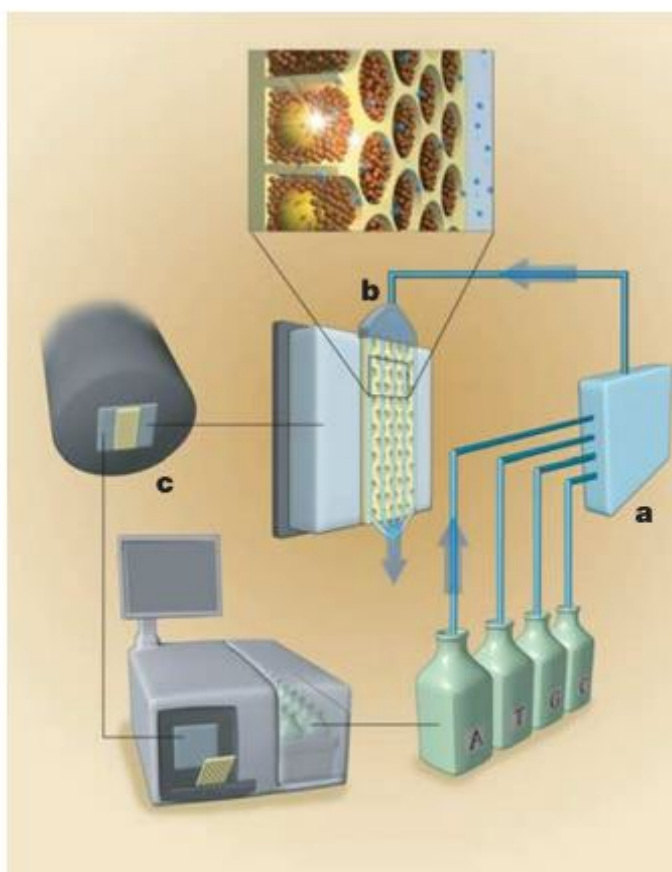


Figure 2 | Sequencing instrument. The sequencing instrument consists of the following major subsystems: a fluidic assembly (a), a flow chamber that includes the well-containing fibre-optic slide (b), a CCD camera-based imaging assembly (c), and a computer that provides the necessary user interface and instrument control.

EX1070.0002.

87. The fourth stage (D) occurred at the same time as complementary strand elongation occurred. As the reactions in the wells emitted light, the 454 sequencer

would detect the intensity of the light and generate intensity graphs for each well. These graphs could then be used in conjunction with software programs to identify the nucleotides in the strand, allowing for researchers to sequence each fragment. The figure below is an example graph.

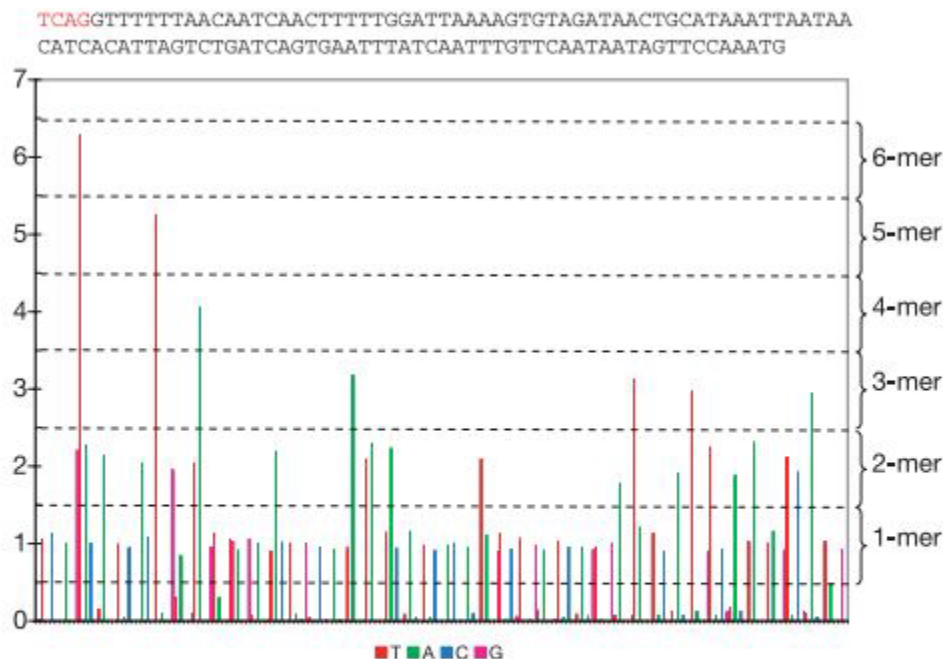


Figure 3 | Flowgram of a 113-bases read from an *M. genitalium* run. Nucleotides are flowed in the order T, A, C, G. The sequence is shown above the flowgram. The signal value intervals corresponding to the various homopolymers are indicated on the right. The first four bases (in red, above the flowgram) constitute the 'key' sequence, used to identify wells containing a DNA-carrying bead.

EX1070.0003.

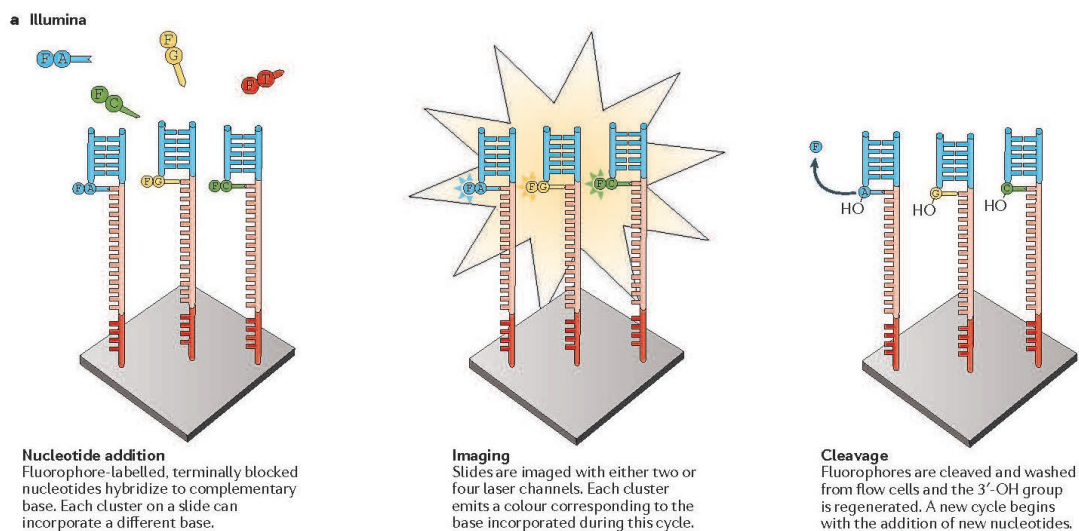
88. The 454 Sequencer exemplified many of the early problems with NGS platforms. The 454 Sequencer relied on the use of light sensors to detect reactions specific to a given nucleotide in order to read sequences. This method, however, was notorious for its high sequencing error rate. EX2003.0001; *see also*

EX2005.0005, 07-08. This issue was particularly pronounced when sequencing homopolymer regions (large stretches of repeating mononucleotides). EX2006.0003; *see also* EX2007.0004. In instances where multiple consecutive identical nucleotides were present, multiple nucleotides could erroneously be incorporated into the reaction at once (rather than just one), increasing the corresponding light signal generated, which affected the calibration of the sequencer, resulting in sequence misreads. EX2003.0009. Additionally, the platform design of measuring light from individual reactions in real time limited its throughput capacity. *Id.*, .0005, 09. Due to its highly specialized design, the platform required technical expertise and expensive reagents and components. EX2008.0002. Further, the error rates and patterning characteristics of the platform necessitated the development of specialized data analysis techniques that relied on specific mathematical algorithms. EX2006.0003. These issues ultimately led to Illumina overtaking the NGS market with Roche announcing the discontinuation of the 454 Sequencer in 2013 since they could no longer compete commercially given the technical limitations of the system. EX2009.0001.

89. Unlike the 454 and Ion Torrent sequence, Illumina sequencers employed a “cyclic reversible termination” sequencing method. EX2003.0003. Samples to be sequenced were first prepared to construct a sequencing library. As with the 454 sequencer, library preparation varied but generally involved the

fragmentation of genetic material followed by adapter ligation and PCR to introduce Illumina platform-specific adapters into the sample. Using these adapters, sequencing library was then attached to a flat, solid surface on the Illumina sequencing instrument via hybridization of the adapters to short complementary sequences located on the sequencer surface. *Id.*, .0004. The attached sequencing library was then amplified on the surface of the sequence to generate copies of the sequencing library. *Id.* Following cluster generation, sequencing commenced. Sequencing began by priming the nucleotide sequencings using universal Illumina primers so that nucleotides could be incorporated into a recently generated complementary stand. *Id.*, .0003. Sequencing occurred in a series of cycles. *Id.* Unlike single-nucleotide addition sequencing approaches, all four nucleotides were provide for each cycle such that any of the four could be incorporated into the elongating complementary strands. *Id.*, .0003-04. Each of the four nucleotide species was labeled with a different fluorescent label. *Id.* Additionally, all of the nucleotides were chemically modified such that no nucleotide could be added after incorporation of a previous chemically modified nucleotide. *Id.* As such, in each cycle only one of the four chemically modified nucleotides were incorporated into each elongating strand. Residual unbound nucleotides were then washed away and the fluorescent signal of each chemically modified nucleotide (reflecting the identity of the nucleotide) was recorded. *Id.* The fluorescent label and chemical modification

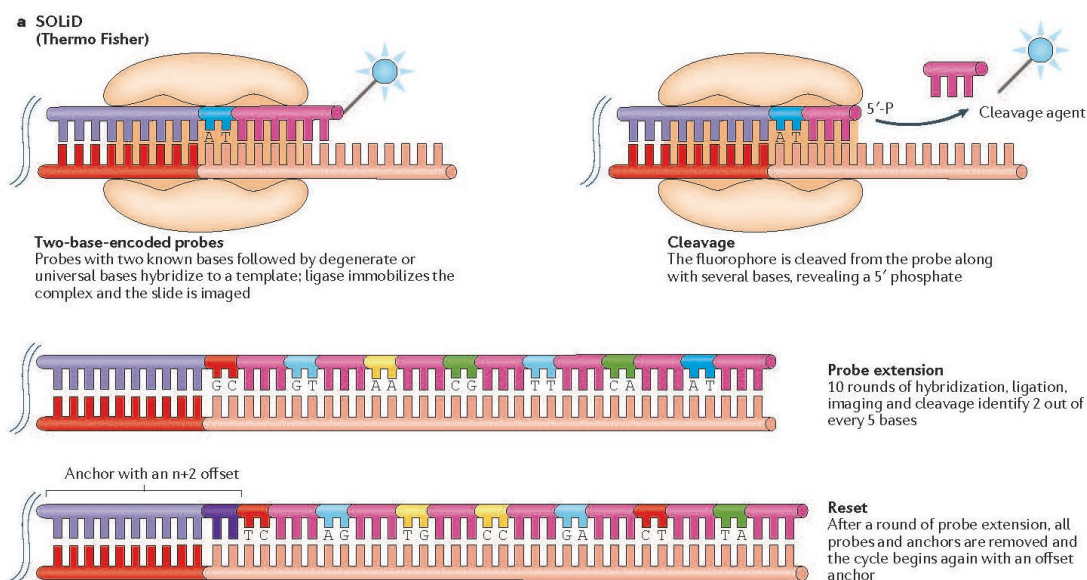
were then removed, allowing for another nucleotide to be added to the nascent complementary strands, and the cycle was repeated. *Id.* Using this process, the identify of each nucleotide incorporated into the complementary strands of the sequencing library could be recorded, and the sequence of the sequencing library could be determined. A figure representing the Illumina sequencing process is provided below.



EX2003.0008.

90. Unlike the Illumina and 454 platforms, the SOLiD sequencer platform utilized a “sequence by ligation” approach. *Id.*, .0003. Library preparation for the SOLiD sequencer was similar to the 454 sequencer, including introduction of adapter sequences, attaching the sequencing library to beads, and emulsion PCR. *Id.*, .0004. Following these steps, prepared libraries would then be sequenced. SOLiD employed a sequence by ligation method wherein labeled probes containing a two

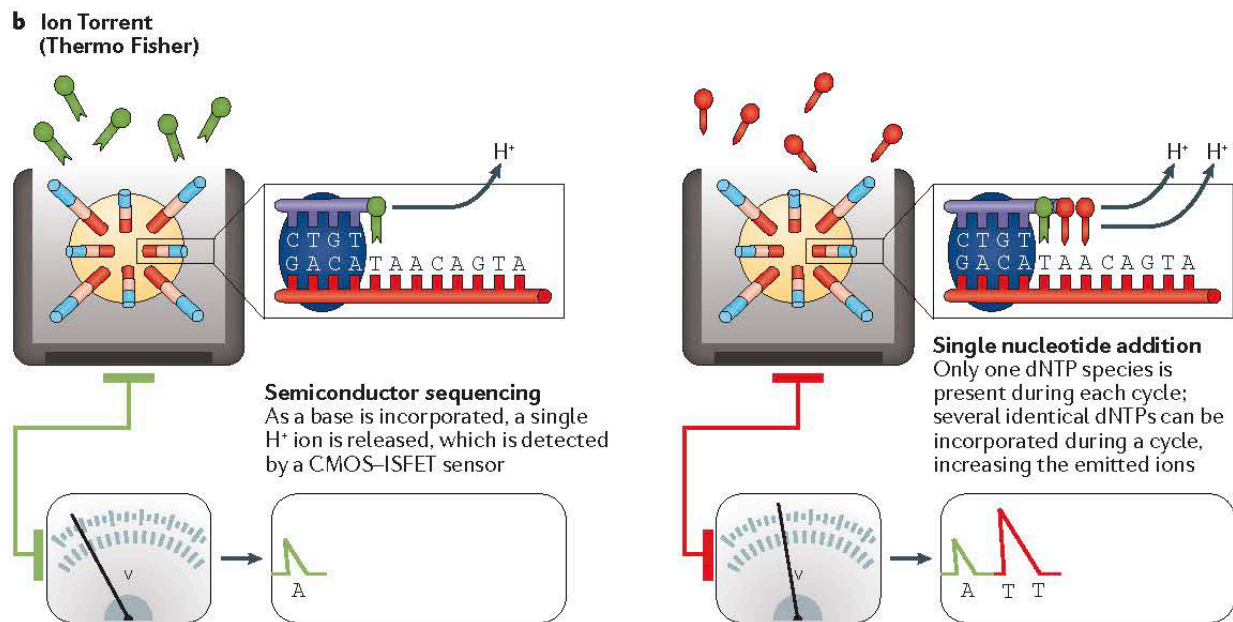
known nucleotides followed by a series of universal nucleotides were hybridized and then ligated to an extending complementary strand corresponding to a nucleotide sequence in the bead-attached sequencing library. *Id.* Each probe was associated with a fluorescent signal. *Id.* After ligation, the fluorescent signal would be recorded, after which the universal nucleotides would be removed from the extending strand so that a subsequent probe could be hybridized and ligated. *Id.* Not all of the universal probe would be removed, so SOLiD sequencing runs required a second sequencing of the same reads using probes that were offset from the initial read through in order to capture the nucleotides bound to the unremoved universal bases, as shown in the below figure. *Id.*, .0006. The SOLiD system used a dinucleotide system wherein each probe contained two nucleotides, leading to 16 possible nucleotide combinations. *Id.*, .0004. The system used only 4 fluorescent signals, meaning that each fluorophore represented one of several possible nucleotide combinations. *Id.* As such, the resulting sequencing data generated by recording the fluorescent signals had to be deconvoluted to identify the sequencing of the corresponding nucleic acid. *Id.* A figure representing the SOLiD sequencing process is provided below.



EX2003.0006.

91. Commercialized after the 454, Illumina, and SOLiD sequencers, the Ion Torrent sequencer utilized a single-nucleotide addition sequencing approach similar to the 454 sequencer. Like the 454 and SOLiD sequencers, Ion Torrent employed similar library preparation methods including loading sequencing libraries onto beads and emulsion PCR. *Id.*, .0004. Ion Torrent also employed a single nucleotide extension approach, wherein a single species of nucleotide would be washed across a bead-containing well plate at a time. *Id.*, .0005. As with the 454 sequencer, the washes were iterative. *Id.* Unlike the 454 sequencer, however, Ion Torrent sequences did not use an optical sensor, and instead employed a pH sensor that detected the change in pH that occurs when a nucleotide is incorporated into a complementary elongating nucleotide strand. *Id.* The pH sensor was capable of imprecise proportional measurements of the intensity of the pH change, “allowing

for limited accuracy in measuring homopolymer lengths.” *Id.*, EX2086.0001-02, EX2098.0008. A figure representing the Ion Torrent sequencing process is provided below.



EX2003.0009.

92. Even before January 2012, Illumina dominated the market. *See* EX2060.0002 (“As of this writing, Illumina dominates the market with 60% of the second-generation sequencer installations, while Life Technologies’ SOLiD system and Roche split nearly all of the remaining market at 19% each.”). J.P. Morgan estimated that while 454 sequencers made up almost 57% of the NGS market in 2007, it was expected to decline to less than 18% by 2012 as Illumina and Ion Torrent became more popular. EX2076.0042. One market survey around the time reported higher consumer dissatisfaction with the 454 sequencers. I have included relevant

consumer responses below, which reflect some of the most salient issues with the 454 sequencer:

- “For Illumina the sample prep is very easy. For 454/AB the prep is a bit more difficult (emulsion PCR step).”
- “The sample prep [for] Illumina is pretty straightforward. A computational graduate student with very little experience got it to work the first time. The 454 and SOLiD instruments both use emulsion PCR, which is a disadvantage.”
- “454 Titanium Emulsion PCR is a problem.”
- “Roche 454 requires a lot of DNA, which means we’re limited in some of the experiments we’d like to do.”

Id., .0019. According to the survey, there was a “prevailing perception” “that Illumina’s sample prep capabilities remain superior.” *Id.*, .0018. When asked to consider the technical problems of the 454 sequencers, respondents identified homopolymers, sequencing speed and cost, and low throughput. *Id.*, .0025. When asked the same for Illumina, many respondents focused on narrower issues and one respondent stated that they were “not aware of any.” *Id.* In a side-by-side comparison of the 454 GS-Flx sequencer, Illumina GAIIx, and SOLiD 3 Plus, the 454 sequencer received the lowest average rating, scoring lower for data accuracy, overall costs, sequencing throughput, ease of use, breadth of applications, and

required sample input volume. *Id.*, .0021. When asked what their greatest concerns were, respondents ranked data accuracy as the highest, followed by the overall cost to sequence and the cost of reagents. *Id.*, .0020.

93. While these second generation NGS platforms were viewed as an overall improvement to sequencing technology, these platforms still failed to adequately address the concerns in the art at the time. For example, these NGS platforms exhibited a higher error rate than Sanger Sequencing per read. EX2003.0001. Additionally, compared to traditional methods, many NGS methods generated significantly shorter individual sequencing reads, while long-read sequencing approaches remained expensive and resource intensive. *Id.* In addition to these limitations, the sheer volume of genetic material that could now be sequenced exacerbated issues with being able to isolate specific regions of genetic material for analysis. EX2004.0003-04. To resolve this issue (among other issues), methods were developed to narrow the window of genomic assessment, such as through “targeted enrichment.”

H. Target Enrichment

94. The inherent size and complexity of the genome has presented longstanding issues when attempting genetic sequencing. Often, researchers and clinicians are interested in select regions of the genome, such as those known to be associated with particular diseases or looking to compare a tumor DNA to normal

DNA to detect changes in specific genes, that are minute in size compared to the genome as a whole. Sequencing methods can struggle to sequence smaller regions of interest in an efficient manner while ensuring sufficient signal (region of interest) to noise (signal from other genomic regions) ratios in the sequencing data and minimizing off-target effects (reactions at other regions of the genome outside the area of interest), all of which are necessary to generate accurate sequencing data.

95. In order to minimize these issues, NGS pipelines were integrated with targeted enrichment approaches designed to enrich a genomic region of interest relative to the rest of the genome prior to sequencing. As understood in the art, the term “enrichment” means to selectively increase the proportion of one part of a sample or reaction mixture over other portions of the sample or reaction mixture. *See, e.g.*, EX2020; EX2010 (describing the use of various prior art target enrichment methods to enrich for target regions). Targeted enrichment methods sought to amplify the nucleotides corresponding to the region of interest while removing or minimizing those reflecting unwanted regions of the genome, thereby enriching the amount corresponding to the signal of the region of interest relative to the rest of the genomic sample.

96. As of the priority date of the '012 and '108 patents, three major categories of target enrichment methods existed: polymerase chain reaction (PCR)-

based methods, capture-by-hybridization, and capture-by-circularization. *See* EX1001.0016 (1:39-46). Each is discussed in further detail below.

1. PCR-based Target Enrichment Methods

97. PCR-based target enrichment methods, as the name suggests, employed PCR to enrich regions of interest from a nucleotide sample prior to sequencing. *See* EX1001.0016 (1:47-50). Primer pairs were designed to target a specific region in a nucleotide sample such that the region was amplified by PCR, while nucleotide sequences outside of the region defined by the primer pair were not amplified. The method could be scaled combining groups of primer pairs that targeted discrete regions to amplify multiple target regions simultaneously, a method termed “multiplex PCR.” *Id.* (1:51-53). Multiplex PCR was useful for targeting small-to-medium-sized genomic regions. PCR-based enrichment was often quicker and more cost-effective than hybridization-based capture approaches but was generally suitable for smaller genomic regions due to limitations in multiplexing capacity and primer design. *See id.*

98. PCR approaches required that sequences within or near the region of interest were known such that primer pairs could be designed to target the region, often precluding true exploratory studies. *See id.* Additionally, primers were often not 100% specific to the primer target sequence (the exact complementary sequence of the primer) and could bind to genomic or complimentary DNA with similar, but

not exact, sequence homology. *See id.* (1:51-53). When sequencing regions of a genome, this issue led to off-target binding to similar sequences outside the region of interest, both generating confounding sequence products that made sequencing interpretation difficult (or impossible) and sequestering primers that would otherwise bind to the actual target sequence. Further, because both the genome and transcriptome are characterized by frequent repetitive sequences, primers designed to target a sequence in a region of interest could inadvertently bind to the same sequence elsewhere in the genome, leading to further off-target binding. 2011.0002. As such, primer-based enrichment methods were often characterized by low specificity and high sequencing background noise.

99. One designer of multiplex PCR kits at the time, QIAGEN, published prior art handbooks discussing these concerns. For example, in the October 2010 QIAGEN Multiplex PCR Handbook, QIAGEN explained that “PCR is an exquisitely sensitive technique that can be used to amplify and detect even a single copy of a nucleic acid sequence,” but that “amplification of such a low number of target sequences is often limited by the generation of nonspecific PCR products and primer-dimers.” EX2089.0037.

100. In addition to the specificity issues discussed above, PCR-based target enrichment methods suffered from design issues that limited their scalability. *See* EX1001.0016 (1:51-53). PCR-based target enrichment methods employed highly

parallel PCR amplification, which required that each target sequence corresponded to a pair of unique, sequence-specific primers. *See id.* (1:48-50). At the time, researchers recognized that “primer design is a critical step in all kinds of RT-PCR methods to guarantee specificity and efficiency of a target amplicon.” EX2090.0001.

101. The prior art at the time taught that it was necessary to use primer pairs rather than a single target specific primer. For example, in 1999, Matz reported issues with “amplification produced by a single primer” and proposed PCR suppression methods to minimize such amplification. EX2091.0001. As another example, in 2008, Bru reported that a single primer mismatch could “greatly decrease PCR efficiency” and taught that “the use of multiple sets of primers targeting different subgroups may be preferable when the sequence polymorphism of the targeted gene family is high.” EX2092.0004. Unfavorable sequencing outcomes using single target-specific primers led researchers to outright reject their use, including one research group that wrote:

While the specificity of individual primers may be testing using BLAST, no further consideration may be given to the influence of non-target sites that can result in suboptimal binding of the primers. ***Accurate quantification requires primer sets that facilitate maximum amplification efficiency.*** In our experience, it is usually necessary to design, synthesize, and validate several primer pairs, until a set is obtained that generates no primer dimers and results in near 100% amplification efficiency.

EX2093.0002 (emphasis added). Other researchers turned to other techniques entirely, such as hybrid capture, to avoid the use of PCR primer pairs. *See, e.g.*, EX2094.0005 (“Unlike PCR-based enrichment methods where both the 3' and 5' ends of targeted DNA must be specified through the use of template-specific primers, hybrid capture allows for the enrichment of sequences containing only partial homology to the viral insertion site.”). Before the claimed inventions, the widely held view in the field was that primer pairs were required to facilitate effective enrichment through amplification, especially PCR amplification. In fact, I am aware of no instances before the priority date in which one target-specific primer was used in connection with PCR amplification to enrich for target sequences.

102. While possible to pool primer pairs together, multiplexing faced several issues. For example, different primer pairs sometimes required different amplification parameters (i.e., temperature and duration settings that facilitate efficient amplification). *See* EX1001.0016 (1:51-53). These diverse amplification parameters were difficult, if not impossible, to uniformly account for in a given PCR as the amplification conditions were often set and held consistent throughout the PCR cycles. Additionally, pooling multiple primer pairs in the same PCR reaction compounded the issues with primer specificity, as previously discussed. Similarly, pooling multiple samples could lower primer sensitivity on a per-sample basis, as multiple samples competed for primer pairs. Moreover, the introduction of more

and varied primers exacerbated the issue of primer dimers, a phenomenon wherein primers bind and amplify each other rather than the true primer target sequence, generating unwanted PCR byproducts and sequestering primers, thereby lowering amplification efficiency. EX2012.0001.

103. In part to address these limitations of multiplex target enrichment by PCR, researchers developed an enrichment method known as “microdroplet-based PCR enrichment” (“microdroplet PCR”) which could be performed using microfluidic devices. *See* EX1001.0016 (1:53-58). Microdroplet PCR required the generation of small solution droplets that contained discrete primer pairs designed for the genomic regions of interest. *Id.* Fragmented DNA samples and PCR reaction mix components could then be introduced into the droplets via a microfluidic chip, which maintained the individual integrity of the primer droplets, generating thousands of discrete PCR reactions while maintaining primer-pair separation. Following targeted enrichment, the droplets could be broken to mix the enriched products into a general solution for sequencing.

104. Microdroplet PCR demonstrated comparable accuracy and coverage to singleplex PCR sequencing while limiting issues with primer specificity and primer-primer interactions characteristic of other multiplex PCR-based target enrichment methods. However, microdroplet PCR required highly sophisticated, specific

instrumentation to run. *See id.* (1:58-63). Further, microdroplet PCR required large numbers of primer pairs when enriching for multiple regions of interest. *Id.*

105. Similar methods were proposed where “several dozen primer pairs can be used together effectively in a multiplexing reaction” by running the reactions in parallel or through microfluidics PCR chips that would allow several thousand singleplex PCRs in parallel. EX2095.0007.

106. Other researchers sought ways to optimize pre-existing multiplex PCR methods by optimizing primer pairs. For example, Shen reported a primer design program called MPprimer, which was designed to address issues with primer specificity for primer pairs. EX2096.0001. Researchers for IDT even published on a “prerequisite for successful multiplex PCR [being] the design of optimal primer pairs,” offering publications to help consumers design optimal primer pairs. EX2097.0005.

2. Capture-by-Hybridization

107. The second major category of enrichment methods was capture-by-hybridization. *See* EX1001.0016 (1:39-46). Like the PCR-based methods discussed above, capture-by-hybridization sought to generate samples containing primarily the genomic region of interest while limiting the presence of extraneous genetic information. Unlike the PCR-based methods, which seek to amplify the signal of the region of interest by copying the region of interest, capture-by-hybridization

methods utilize long oligonucleotide chains, termed “probes,” that are designed to hybridize directly to sequences that define, overlap substantially with, or are in close proximity to the region of interest. *See id.* (1:64-66). The probes could then be selectively isolated, allowing the genomic regions of interest to be physically separated from the unwanted genetic material. *See id.* (1:64-2:4).

108. As capture-by-hybridization relied on target-specific probes to sequester targets for sequencing, capture-by-hybridization was less susceptible to background noise generated by off-target sequencing as compared to PCR-enrichment methods. Capture-by-hybridization approaches required larger amounts of sample input than PCR-based methods. *See id.* (2:4-10). Additionally, capture-by-hybridization required complex probes designed to bind the region of interest, and the capture and selection design necessitated lengthy and complicated workflows as compared to PCR-based methods. *Id.* While the probe design of capture-by-hybridization allowed for theoretically large-scale amplicon panels, the size of capture-by-hybridization panels was limited by the higher costs associated with individual nucleotide probes. *Id.* Additionally, capture-by-hybridization methods were characterized as biased towards repetitive sequence elements present in the genome, which could lead to their over-representation in corresponding sequencing data. EX2013.0001.

3. Capture-by-Circularization

109. The third major category of enrichment methods was capture-by-circularization. *See* EX1001.0016 (1:39-46). Capture-by-circularization methods utilized linear nucleotide probes designed such that both ends of the probes hybridized at or near adjacent regions in a genomic region of interest. *See Id.* (2:11-14). When properly hybridized, the region between the hybridized regions bound by both ends of the probe could then be copied into the probe while joining the two ends of the probe to create a circle, a process termed “gap fill polymerization.” *Id.* (2:14-16). Probe circularization allowed for the selective enrichment of circularized probes by filtering out linear probes and probes that did not circularize properly. Following selection, circularized probes could be re-linearized and the copied region of interest could be further amplified by PCR. *See id.* (2:16-18).

110. Due to the multiple hybridization approach, capture-by-circularization was highly specific for the target region of interest. However, capture-by-circularization probes exhibited a lack of consistent binding to the desired target region of interest, resulting in a variable amount of sequence coverage. *See id.* (2:18-24). Further, the large size of the probes limited the amount of sequencing real estate for the sequences of interest in a given sequencing read, as regions of the probes themselves often took up a sizable portion of the generated sequencing read.

111. As late as 2011, the art recognized that these three target capture approaches suffered from issues of varying degree with “uniformity, efficiency of coverage and off-target capture.” EX1050.0007. Given these issues, practitioners wondered whether “it may ultimately be more cost-effective to sequence a whole genome” rather than use target capture. *Id.*

V. THE PERSON OF ORDINARY SKILL IN THE ART

112. The Board adopted the following definition in its Institution Decision:

[A] POSA would have had a Ph.D. in the field of chemistry, molecular biology, or biochemistry. Alternatively, a POSA would have had a B.S. or M.S. degree with three to five years of experience working in any of those fields. Additionally, the POSA would have had experience with NGS sequencing methods and a working knowledge of how those methods functioned.

Paper 9 (IPR2025-00015), 14; Paper 7 (IPR2025-00016), 14.

113. I agree with the Board that a POSA would possess either a Ph.D. in the field of chemistry, molecular biology, or biochemistry; or a B.S. or M.S. degree and approximately three-to-five years of experience working in a molecular biology laboratory. However, it is my opinion that a POSA would have only to have routine experience with DNA sequencing methods (including next-generation sequencing which was the dominant mode at the time) and a working knowledge of how these sequencing methods worked in order to troubleshoot experiments—not experience and a working knowledge specific to NGS sequencing.

114. For the purposes of this Declaration, I apply the Board’s definition of a POSA. Slight variations on this definition of a POSA would not change my opinions set forth in this declaration. Under this definition of a POSA, I possessed the relevant qualifications as of January 26, 2012. *See supra* ¶¶ 12-21. My views reflect those of a POSA at the time.

115. Additionally, I have reviewed the qualifications and background of Dr. Eli Meyer. Under this definition of a POSA, Dr. Meyer possessed the relevant qualifications as of January 26, 2012. *See* EX2287.¶¶ 5-11. His views reflect those of a POSA at the time.

VI. THE PATENTED INVENTIONS

116. Titled “Compositions and Methods for Targeted Nucleic Acid Sequence Enrichment and High Efficiency Library Generation,” the ’012 and ’108 patents are broadly directed to “methods, compositions and kits for targeted nucleic acid sequence enrichment in a nucleic acid sample and for high efficiency nucleic acid library generation for next generation sequencing (NGS).” EX1001.0001; EX1002.0001. The patents are specifically directed to methods “for the selective enrichment of a plurality of defined target sequences from complex DNA with a set of common primers and adaptors, thus circumventing the need for multiplex PCR and multiple primer pairs.” EX1001.0018 (6:12-16); EX1002.0018 (6:14-18). The methods improve on the prior art by creating “a simple, low cost, high throughput

system for target enrichment and library preparation.” EX1001.0018 (6:25-27); EX1002.0018 (6:27-29).

A. The Claims of the '012 Patent

117. The claimed inventions of the '012 patent are defined in 22 claims. The sole independent claim of the '012 patent, claim 1, is reproduced below:

Claim Element	Claim Language
1[pre]	A method for sequencing an enriched nucleic acid sequence of interest, the method comprising:
1[a]	obtaining a nucleic acid fragment ligated to a partial duplex adaptor, wherein the nucleic acid fragment comprises a nucleic acid sequence of interest, wherein the partial duplex adaptor comprises a first adaptor sequence, and wherein the partial duplex adaptor comprises a first strand and a second strand, wherein the first strand is longer than the second strand;
1[b]	annealing one or more oligonucleotides in solution to the nucleic acid sequence of interest in the nucleic acid fragment ligated to the partial duplex adaptor, wherein the one or more oligonucleotides comprise a 3' portion with at least 10 bases designed to be complementary to the nucleic acid sequence of interest and a 5' tail portion comprising a second adaptor sequence that is non-complementary to the nucleic acid sequence of interest;

Claim Element	Claim Language
1[c]	extending the one or more oligonucleotides annealed to the nucleic acid sequence of interest in the nucleic acid fragment ligated to the partial duplex adaptor with a polymerase, thereby generating one or more oligonucleotide extension products comprising sequence complementary to the first adaptor sequence at a first end, sequence complementary to the nucleic acid sequence of interest, and the second adaptor sequence at a second end;
1[d]	amplifying the one or more oligonucleotide extension products using a first primer that anneals to a complement of the first adaptor sequence and a second primer that anneals at its 3' end to a complement of the second adaptor sequence to enrich for the nucleic acid sequence of interest, wherein products of the amplifying comprise a 3' end with sequence complementary to a sequence on a surface;
1[e]	annealing a strand of the products of the amplifying to the sequence on the surface using the 3' end with sequence complementary to the sequence on the surface; and
1[f]	sequencing the enriched nucleic acid sequence of interest on a massively parallel sequencing platform.

118. The '012 patent contains 21 dependent claims, which are reproduced below:

Claim	Claim Language
2	The method of claim 1, wherein the nucleic acid fragment comprises genomic DNA.
3	The method of claim 1, wherein the nucleic acid fragment comprises cDNA.
4	The method of claim 1, wherein the nucleic acid fragment comprises RNA.
5	The method of claim 1, wherein the first adaptor sequence and the second adaptor sequence are distinct from each other.
6	The method of claim 1, wherein the one or more oligonucleotides comprise more than one oligonucleotide, and wherein the second adaptor sequence is common to the more than one oligonucleotides.
7	The method of claim 1, wherein the first adaptor sequence, the second adaptor sequence, or both comprise barcode sequence.
8	The method of claim 7, wherein the barcode sequence is at least 8 nucleotides in length.
9	The method of claim 7, wherein the barcode sequence is 12 nucleotides in length.
10	The method of claim 1, wherein the obtaining comprises obtaining a plurality of nucleic acid fragments ligated to a partial duplex adaptor, wherein each partial duplex adaptor of each nucleic acid fragment of the plurality of nucleic acid fragments comprises a barcode, and wherein each barcode is unique.
11	The method of claim 1, wherein the nucleic acid fragment ligated to the partial duplex adaptor is double-stranded prior to step b).
12	The method of claim 11, further comprising denaturing the nucleic acid fragment ligated to the partial duplex adaptor prior to step b), thereby generating a single-stranded nucleic acid fragment

Claim	Claim Language
	comprising the nucleic acid sequence of interest ligated to the first adaptor sequence.
13	The method of claim 1, wherein the polymerase is a DNA polymerase.
14	The method of claim 1, wherein the first strand of the partial duplex adaptor is ligated to a 5' end of the nucleic acid fragment.
15	The method of claim 1, wherein the second strand of the partial duplex adaptor is ligated to a 3' end of the nucleic acid fragment.
16	The method of claim 1, further comprising fragmenting a nucleic acid, thereby generating a nucleic acid fragment; and purifying the nucleic acid fragment prior to obtaining the nucleic acid fragment ligated to the partial duplex adaptor in step a).
17	The method of claim 1, further comprising performing bridge amplification of the annealed strand after step e).
18	The method of claim 1, wherein the sequencing comprises use of four labeled reversible terminators.
19	The method of claim 1, wherein the sequencing comprises semiconductor sequencing.
20	The method of claim 1, further comprising fragmenting a nucleic acid, thereby generating a nucleic acid fragment; and performing end repair on the nucleic acid fragment prior to obtaining the nucleic acid fragment ligated to the partial duplex adaptor in step a).
21	The method of claim 1, wherein the nucleic acid fragment of the nucleic acid fragment ligated to the partial duplex adaptor in step a) is obtained by performing reverse transcription on an RNA fragment prior to obtaining the nucleic acid fragment ligated to the partial duplex adaptor in step a).
22	The method of claim 1, wherein each end of the nucleic acid fragment is ligated to a partial duplex adaptor.

B. The Claims of the '108 Patent

119. The claimed inventions of the '108 patent are defined in 19 claims. The sole independent claim of the '108 patent, claim 1, is reproduced below:

Claim Element	Claim Language
1[pre]	A method for sequencing an enriched nucleic acid sequence of interest, the method comprising:
1[a]	annealing one or more oligonucleotides in solution in a reaction mixture to the nucleic acid sequence of interest in a nucleic acid fragment, wherein the reaction mixture comprises a plurality of nucleic acid fragments, wherein the nucleic acid fragment comprising the nucleic acid sequence of interest comprises a first adaptor sequence, wherein the one or more oligonucleotides comprise a 3' portion with at least 10 bases designed to be complementary to the nucleic acid sequence of interest and a 5' tail portion comprising a second adaptor sequence that is non-complementary to the nucleic acid sequence of interest;
1[b]	extending the one or more oligonucleotides annealed to the nucleic acid sequence of interest in the nucleic acid fragment comprising the first adaptor sequence with a polymerase, in the reaction mixture, thereby generating one or more oligonucleotide extension products comprising sequence complementary to the first adaptor sequence at a first end, sequence complementary to the nucleic acid sequence of interest, and the second adaptor sequence at a second end;

Claim Element	Claim Language
1[c]	amplifying the one or more oligonucleotide extension products, in the reaction mixture, using a first primer that anneals to the complement of the first adaptor sequence and a second primer that anneals at its 3' end to a complement of the second adaptor sequence, thereby enriching the nucleic acid sequence of interest, by generating amplified products comprising the enriched nucleic acid sequence of interest; and
1[d]	sequencing the amplified products comprising the enriched nucleic acid sequence of interest on a massively parallel sequencing platform.

120. The '108 patent contains 18 dependent claims, which are reproduced below:

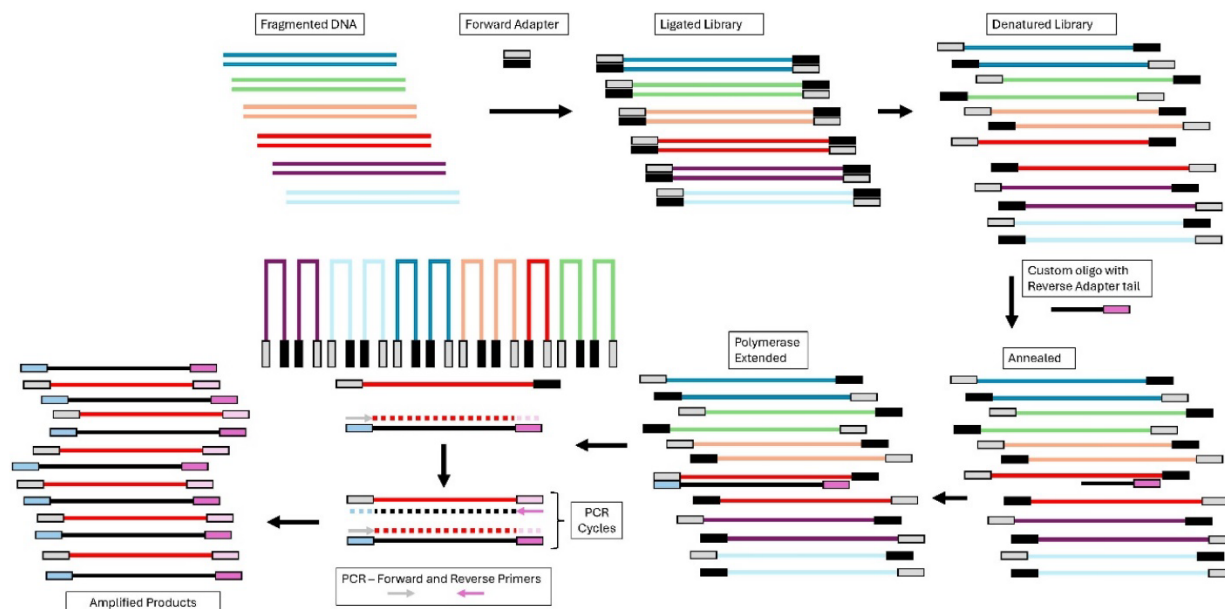
Claim	Claim Language
2	The method of claim 1, wherein the nucleic acid sequence of interest comprises genomic DNA.
3	The method of claim 1, wherein the nucleic acid sequence of interest comprises cDNA.
4	The method of claim 1, wherein the nucleic acid sequence of interest comprises RNA.
5	The method of claim 1, wherein the first adaptor sequence and the second adaptor sequence are distinct from each other.
6	The method of claim 1, wherein the plurality of nucleic acid fragments comprises the first adaptor sequence, and wherein the first

Claim	Claim Language
	adaptor sequence is common to the plurality of nucleic acid fragments.
7	The method of claim 1, wherein the one or more oligonucleotides comprises more than one oligonucleotide, wherein the second adaptor sequence is common to the more than one oligonucleotide.
8	The method of claim 1, wherein the first adaptor sequence and/or the second adaptor sequence comprise a barcode sequence.
9	The method of claim 1, further comprising denaturing double-stranded nucleic acid fragments prior to step a).
10	The method of claim 1, wherein the polymerase is a DNA polymerase.
11	The method of claim 1, wherein the sequencing comprises the use of four labeled reversible terminators.
12	The method of claim 1, wherein the sequencing comprises semiconductor sequencing.
13	The method of claim 1, wherein the sequencing comprises nanopore sequencing.
14	The method of claim 1, wherein each of the plurality of nucleic acid fragments comprises the first adaptor sequence at a 5' end.
15	The method of claim 1, further comprising fragmenting nucleic acids prior to step a), thereby generating the plurality of nucleic acid fragments.
16	The method of claim 1, further comprising ligating the first adaptor sequence to a nucleic acid fragment in the plurality of nucleic acid fragments prior to step a), thereby generating the nucleic acid fragment comprising the first adaptor sequence.
17	The method of claim 1, wherein the amplified products comprise a 3' end with sequence complementary to a sequence on a surface.

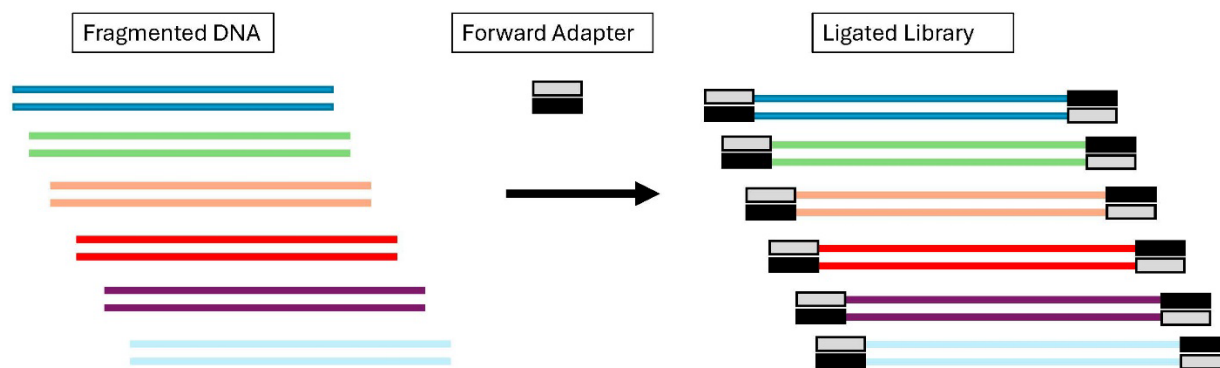
Claim	Claim Language
18	The method of claim 17, further comprising annealing a strand of the products of the amplifying to the sequence on the surface using the 3' end with sequence complementary to the sequence on the surface prior to step d).
19	The method of claim 17, wherein the surface is a surface of a flow cell.

C. Illustration of the Claimed Inventions

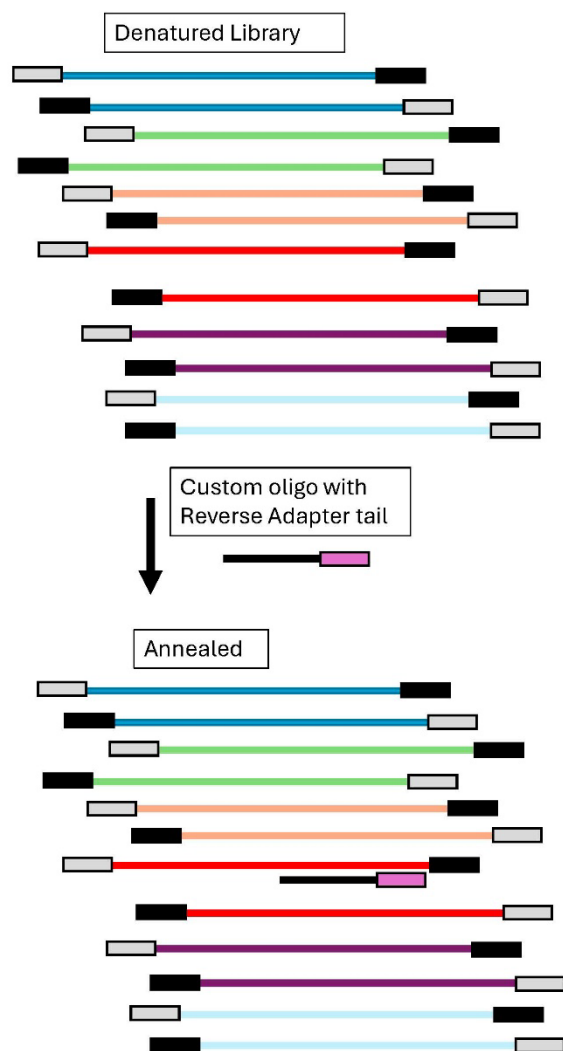
121. I have prepared a series of figures to illustrate how a POSA would understand the claimed limitations. Starting with claim 1 of the '012 patent, a POSA would understand the method is directed to the sequencing of an enriched nucleic acid sequence of interest. The POSA would also understand that many of the claimed steps are directed to library preparation methods for amplicons that are subsequently sequenced with NGS. A representation of the method disclosed in claim 1 is depicted below.



122. In the first step, a POSA would understand that a nucleic acid fragment must be obtained. This nucleic acid fragment must contain a nucleic acid sequence of interest (depicted below as the red fragment). The fragment must also be ligated to at least one partial duplex adaptor (depicted below as the gray and black “forward adapter”), which contains a first adaptor sequence. Ligation can occur at one or both ends of the fragment. While not depicted in the illustration, a POSA would understand that one strand of the partial duplex adaptor is longer than the other, and sits as a single-stranded “overhang.” Because the claim is directed to targeted enrichment of the sequence, a POSA would necessarily understand that this fragment is not alone in the solution, but that instead there are thousands or millions of other fragments in the solution too (the non-red fragments in the depiction).

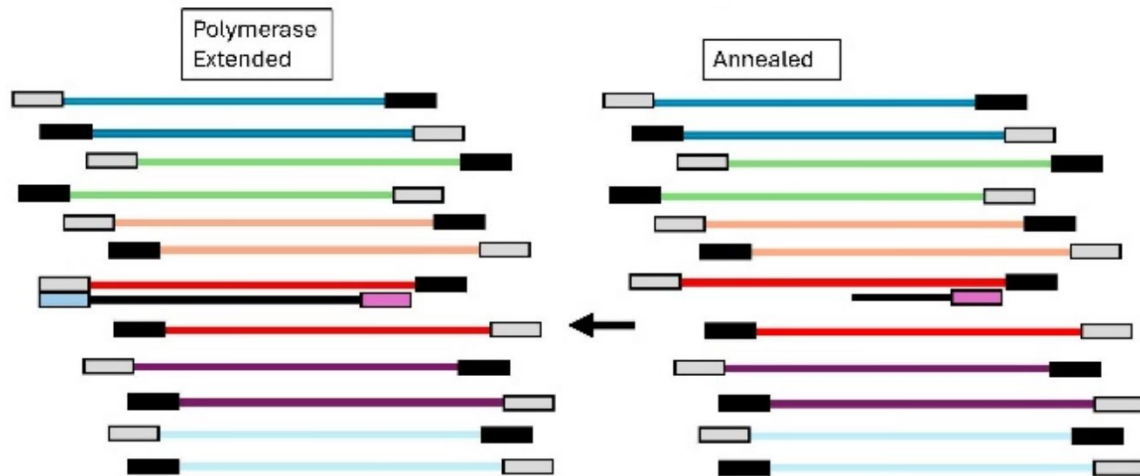


123. In the second step, a POSA would understand that at least one oligonucleotide with sequence complementary to the sequence of interest is annealed to the sequence of interest (depicted below as the “custom oligo with reverse adapter tail”). This oligonucleotide must contain at least 10 bases that are complementary to the sequence of interest (depicted below as the black line). It must also contain a second adaptor sequence that is not complementary to the sequence of interest (depicted below as the pink box).

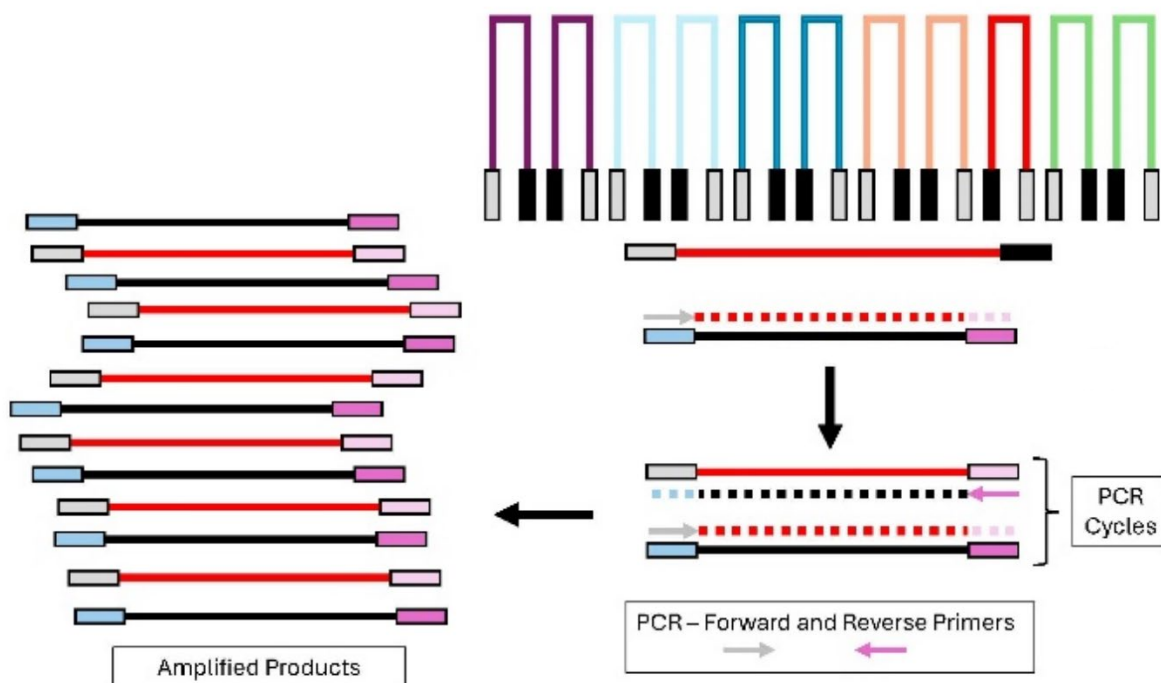


124. In the third step, a POSA would understand that the annealed oligonucleotide is extended, creating an oligonucleotide extension product (depicted below on the left as the longer black strand attached on the left to the blue box and on the right to the pink box). The extension product must contain at least three components: (1) sequence complementary to the first adaptor sequence at one end (depicted below as the blue box); (2) sequence complementary to the nucleic acid sequence of interest (depicted as the black line); and (3) the second adaptor sequence

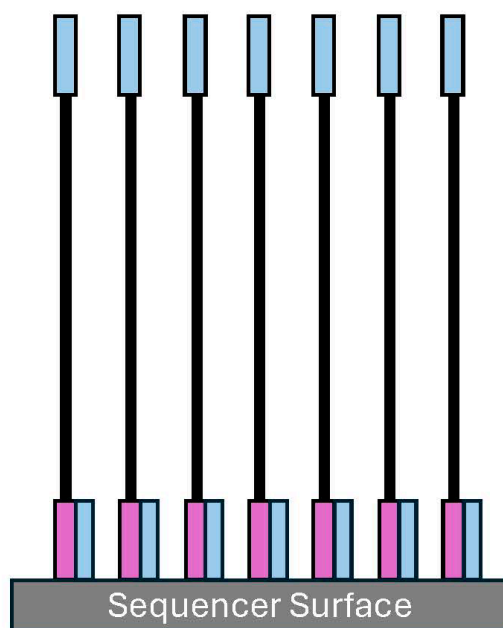
at the other end (depicted below as the pink box). It would be understood that no other fragments but the extension products contain the first and third components.



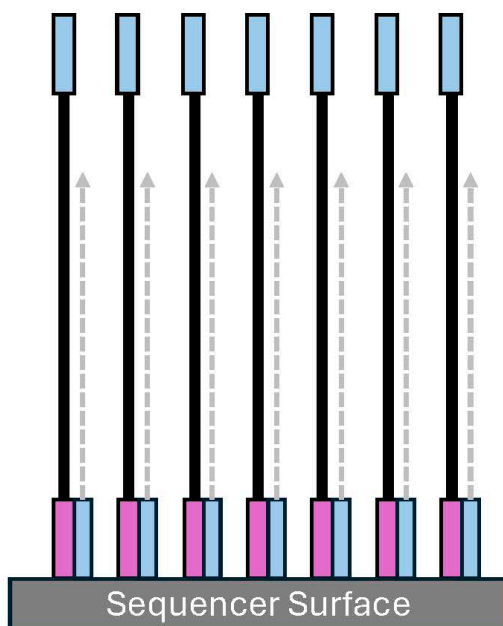
125. In the fourth step, a POSA would understand that the extension products are amplified using two primers: one that anneals to the complement of the first adaptor sequence and another that anneals as its 3' end to a complement of the second adaptor sequence (depicted below as the grey and pink arrows, respectively). Because only extension products with the nucleic acid sequence of interest will have the correct primer binding sites, only the extension products are amplified. This amplification process will necessarily enrich for the sequence of interest by creating a library of amplicons with the sequence of interest and a 3' end with sequence complementary to a sequence on a surface (depicted on the left). Although the figure shows the non-amplified fragments being bent, a POSA would understand that this is meant to show that the fragments are not undergoing amplification and not any specific mechanism, such as PCR suppression.



126. In the fifth step, a POSA would understand that the amplicons are annealed to a surface through their 3' ends. A POSA would also understand that this surface is necessary for fixing the amplicons for sequencing, such as Illumina flow cells.



127. In the sixth and final step, a POSA would understand that the annealed amplicons with the sequence of interest are sequenced with a massively parallel sequencing platform (i.e., NGS).



128. Claim 1 of the '108 patent discloses a similar method. It is also a method for sequencing an enriched nucleic acid sequence of interest. The first step of this method is to anneal the oligonucleotide to a sequence of interest. As Dr. Sims acknowledges, this step is identical to the corresponding annealing step of claim 1 of the '012 patent. EX1085.0170-0171. While the '108 patent specifies that the annealing occurs “in solution in a reaction mixture” wherein the “reaction mixture comprises a plurality of nucleic acid fragments,” a POSA would have understood that the same limitation would necessarily exist for the '012 patent. If the fragment containing the sequence of interest were isolated from fragments not containing the sequence of interest, then it would be impossible to enrich for fragments containing the sequence of interest (i.e., a solution starting with 100% of sequence X cannot be increased beyond 100%).

129. In the second step, the oligonucleotide is extended. Dr. Sims acknowledges that this step is identical to the extension step of claim 1 of the '012 patent. *Id.*, .0172. As with the '012 patent, this step creates oligonucleotide extension products with (1) sequence complementary to the first adaptor sequence at one end; (2) sequence complementary to the nucleic acid sequence of interest; and (3) the second adaptor sequence at the other end. Because the patent uses the two adaptor sequences to amplify and thus enrich for the sequence of interest, no other fragments in the solution can contain both adaptor sequences at their ends.

130. In the third step, the oligonucleotide extension products are amplified to enrich for the sequence of interest. Dr. Sims acknowledges that this step is identical to the amplification step of claim 1 of the '012 patent. *Id.*, .0173. While this step does not expressly require that the amplicons contain sequence complementary to the second adaptor sequence, a POSA would have understood that the amplification of the oligonucleotide extension products would necessarily result in that outcome. Dependent claim 17 specifies that this sequence must be complementary to a sequence on a surface (with dependent claim 18 further specifying that the amplicons are annealed to the sequence on the surface prior to sequencing). Additionally, a POSA would have understood that the enriched nucleic acid sequence of interest is based on the population of fragments that contain the two adaptor sequences, which are used to facilitate sequencing. This is because the purpose of the method is to sequence an enriched nucleic acid sequence of interest. Enriching for the sequence of interest but failing to incorporate adaptor sequences that can facilitate sequencing defeats the purpose of the method. Additionally, the patents disclose that the incorporation of these adaptor sequences should occur prior to the enrichment step so as to facilitate an efficient library preparation method. EX1001.0033, cl. 1; EX1002.0033, cl. 1.

131. In the fourth and final step, the amplicons are sequenced with an NGS platform. Dr. Sims acknowledges that this step is identical to the sequencing step of claim 1 of the '012 patent. EX1085.0174.

D. Relevant Prosecution History

132. I understand that the patents at issue, the '012 Patent and the '108 Patent, are part of a larger patent family containing additional related patents. I understand that the prior assignee of the patents at issue, NuGEN, filed the first application in the patent family, U.S. Provisional Patent Application No. 61/591,241 on January 26, 2012. EX1001.0001; EX1002.0001. I understand that NuGEN subsequently filed U.S. Patent Application No. 13/750,768 (“the '768 application”), on January 25, 2013. *Id.* I understand that U.S. Patent Application No. 13/750,768 issued as U.S. Patent No. 9,650,628 (“the '628 patent”) on May 16, 2017. *See id.*

133. I understand that NuGEN filed U.S. Patent Application No. 15/471,785 (“the '785 application”) as a continuation of the '768 application on March 28, 2017. EX1001.0001. I understand that the '785 application issued as the '012 Patent on July 31, 2018. *Id.*

134. I understand that NuGEN filed U.S. Patent Application No. 16/017,340 (“the '340 application”) as a continuation of the '785 application on June 25, 2018. EX1002.0001. I understand that the '340 application issued as the '108 Patent on December 29, 2020. *Id.*

135. I understand that the '012, '108, and '628 Patents all claim priority to Provisional Application No. 61/591,241, which was filed on January 26, 2012. EX1001.0001; EX1002.0001.³ I further understand that, for purposes of this proceeding, IDT does not dispute that the priority date as reported on the faces of the '012 and '108 Patents is January 26, 2012. Petition, 28.

136. I understand that during prosecution of the applications that would issue as the '012, '108, and '628 patents, the same examiner rejected the proposed claims arguing that they were obvious over two prior art references: U.S. Patent Publication No. 2005/0142577 (“Jones”) and Tucker *et al.* “Massively parallel sequencing: the next big thing in genetic medicine” (“Tucker”).

137. I understand that the Applicant traversed the Examiner’s rejections, arguing that a POSA would not have had a motivation to combine Jones and Tucker to practice the claimed inventions with a reasonable expectation of success. I understand that as part of the Applicant’s responses, the Applicant submitted a sworn affidavit by Dr. Andrew I. Brooks, in which Dr. Brooks presented evidence and testimony concerning why a POSA would not have been motivated to adapt a

³ I understand that the '012 and '108 patents share a common specification. Because of this, I sometimes cite to one of the two specifications herein. To the extent relevant, those references apply equally to the other patent.

method using a hybridization array for NGS and why a POSA would not have had a reasonable expectation of success. I also understand that Dr. Brooks provided evidence and testimony about the same through his accounts of skepticism and unexpected results.

138. Dr. Brooks was widely known and respected in the industry as being at the forefront of genomic sequencing technology. During his decades of work in the field, Dr. Brooks was the Chief Operating Officer of Rutgers University Cell and DNA Repository (“RUCDR”) Infinite Biologics (in addition to being a long distinguished research professor at Rutgers). At the time, RUCDR was known as the world’s largest university-based biorepositories and one of the largest sequencing laboratories in the country. Dr. Brooks was considered an authority in the field for high throughput sample analysis, having authored numerous publications on the topic. His work included various applied applications to genomic analysis, including diagnostic testing. For example, towards the end of his career at RUCDR, Dr. Brooks develop the first FDA-approved rapid saliva test for COVID-19. The New York Times described this invention as “one of the most reliable means of determining whether someone has the coronavirus.” EX2032.0001. It even prompted Governor Phil Murphy of New Jersey to call Dr. Brooks “one of the state’s unsung heroes.”

139. Under either IDT or Tecan's proposed definitions of a POSA, Dr. Brooks would have easily met the criteria and would have been considered an authoritative expert on high-throughput sequencing technology.

140. I understand that one issued raised by IDT is the notion that the Examiner reviewing the '785 application did not consider prior art where the extension step occurred as part of a PCR reaction. Petition, 22-24. I have been asked to review the prior art references cited during prosecution to assess whether any disclose an extension step that occurred as part of a PCR reaction. Based on my review, there are at least two that do: U.S. Patent Publication No. 2003/0232348 ("Jones 2003," provided as EX1039) and U.S. Patent Publication No. 2012/0309002 ("Link 2012," provided as EX2014).

141. The Examiner relied on three specific disclosures from Jones 2003 to argue that it taught the claimed extension step: Figure 3, paragraph 94, and paragraph 129. EX2019.0006. Paragraph 94 discloses that the "capture probes are extended in the presence of polymerase and dNTPs." EX1039.0019, [0094]. As explained above, polymerase and dNTPs are commonly used in PCR to extend a complementary strand. *See supra* ¶¶ 42-44. Nothing in Jones 2003 suggests that PCR would not be performed for extension here. Further, paragraph 129 describes an example ("multiplexed anchored runoff amplification") where Applied Biosystems' AmpliTaq Gold polymerase was used with Taq Gold Buffer, dNTPs,

and TE buffer to extend capture probes. EX1039.0023, [0129]. AmpliTaq Gold polymerase and Taq Gold Buffer were well known components of PCR reactions—including extension through PCR—before the publication of Jones 2003. *See, e.g.*, EX2015.0002; 2016.0003. TE buffer was also a commonly used solution to help protect nucleic acids as they undergo PCR. *See* EX2017.0009-10. Based on the disclosed reagents, I understand Example 1 of Jones 2003 to disclose an extension step that occurred as part of a PCR reaction. I am not aware of another type of extension reaction that Jones would have done using these reagents.

142. Additionally, Jones discloses that for extension, “the reaction was incubated at 95° C for 6 min followed by 26 cycles of 95° C for 30 sec, 68° C for 2.5 min (decreasing 0.5° C on each subsequent cycle) and 72° C for 1 min, then to 4° C.” EX1039.0023, [0129]. This protocol illustrates multiple cycles of denaturing, annealing, and extension, what would be commonly understood to be PCR. *See supra* ¶¶ 42-44.

143. The Examiner also referred to Link 2012 as “pertinent to applicant’s disclosure” in an office action rejecting the proposed claims of the ’768 parent application on July 15, 2015. EX2019.0015-16. Link is a patent publication directed to a “method for preparing a droplet for multiplex chemical reactions.” EX2014.0015. Link discloses that in certain embodiments of the claimed inventions, “the amplification reaction is the polymerase chain reaction.” *Id.*, .0010, [0050].

Link explains that “[t]o effect amplification, primers are annealed to their complementary sequence within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. *The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one cycle; there can be numerous cycles) to obtain a high concentration of an amplified segment of a desired target sequence.*” *Id.* [0051] (modified). This disclosure is of extension occurring as part of a PCR reaction.

E. Claim Construction

144. I understand that IDT has sought the construction of two terms: “sequence of interest” and “barcode.”

145. I understand that the Panel construed “sequence of interest” to mean “a polynucleotide sequence of interest, for which enrichment is desired.” Paper 7, 15. I do not disagree with that construction.

146. For the term “barcode,” I understand that the Panel declined to construe the term since the claims in which it appears were not separately at issue in Patent Owner’s Preliminary Response. I also understand that the status of the disputes surrounding those claims remain the same, so it is unnecessary to construe the term.

147. I understand that the Panel construed the term “enrich” to have its plain and ordinary meaning of “to process so as to add or increase the proportion of a

desirable ingredient.” *Id.* I do not disagree with that construction. However, I believe further construction of the term in the context of the patent claims is necessary.

148. Additionally, for the ’012 patent, I believe it is necessary that the Panel construe “RNA fragments” to have its plain and ordinary meaning.

1. “enrich”

149. The Board construed “enrich” to have its plain and ordinary meaning of “to process so as to add or increase the proportion of a desirable ingredient.” *Id.* I agree that a POSA would have considered the term “enrich” to have its plain and ordinary meaning and that the Panel’s definition reflects that meaning.

150. In the context of the patents, this term is used in a few instances. The first refers to “an enriched nucleic acid sequence of interest.” A POSA would understand that based on the plain and ordinary meaning of “enrich,” a nucleic acid sequence of interest is enriched when the proportion of the nucleic acid fragments containing the sequence of interest is increased relative to other nucleic acid fragments in the sample. A POSA would understand that the proportion of a nucleic acid fragment in a sample can only be increased where there are other fragments in the solution; i.e., if a solution contains 100% of a single fragment type, the proportion of that fragment type cannot be increased any further.

151. I have considered how Dr. Sims uses the concept of enrichment in his papers and have concluded that he applies the term in the same manner. For example, in his declaration, Dr. Sims opines that the claimed “library is selectively enriched for a particular ‘sequence of interest,’ such that the sequence of interest appears in the sequencing library at a higher relative proportion that it appeared in the original sample from which the sequencing library was created.” EX1085.¶ 30. I understand him to mean that the proportion of the fragment containing the sequence of interest is increased relative to the other nucleic acid fragments in the library. At his deposition, Dr. Sims explained that “an enrichment process results in overrepresentation of certain nucleic acid sequences relative to others in a sample.” EX2037.0028 (173:12-17). I again understand him to be referring to the same concept.

152. The patents use the term “enrich” in a second context. For the ’012 patent, the term is used in the step “amplify the one or more oligonucleotide extension products . . . to enrich for the nucleic acid sequence of interest.” For the ’108 patent, the term is used in the step “amplifying the one or more oligonucleotide extension products . . . thereby enriching the nucleic acid sequence of interest, by generating amplified products comprising the enriched nucleic acid sequence of interest.” I understand both patents to use the term “enrich” in the same manner. A POSA would understand that this term as used requires that the claimed

amplification process results in the enrichment of the nucleic acid sequence of interest.

153. As explained above, the amplification step for the patents relies on two primers to make copies. *See supra*, ¶125. The first primer anneals to a complement of the first adaptor sequence. A POSA would understand the complement of the first adaptor sequence is a component generated during the previous step. A POSA would also understand that because only oligonucleotide extension products will have the complement of the first adaptor sequence (other fragments have just the first adaptor sequence through the claimed ligation step), only the oligonucleotide extension product sequence is amplified with this first primer. The second primer anneals at its 3' end to a complement of the second adaptor sequence. Because neither the oligonucleotides with sequence complementary to the nucleic acid sequence of interest nor the oligonucleotide extension products contain sequence complementary to the second adaptor sequence (both contain the second adaptor sequence), a POSA would understand that only the amplified products generated by using the first primer—i.e., the complement to the oligonucleotide extension product sequence—contain the requisite primer binding site. This means that in the claimed process, amplification only occurs for the oligonucleotide extension product sequence and its complement. This limitation is what allows for the amplification to enrich for the nucleic acid sequence of interest.

154. The common specification of the patents supports this understanding. The specification explains that the claimed methods “can be useful to generate . . . libraries with enriched population of sequence regions of interest.” EX1001.0028-29 (26:64-27:2). In describing the claimed amplification step, the specification explains “[i]n this manner only oligonucleotide extension products comprising both the first and second adaptor sequence will be amplified and thus enriched.” *Id.*, .0021 (12:14-17).

155. Figure 1 of the patents illustrates this requirement that the amplification process must enrich for the sequence of interest:

As illustrated in FIG. 1, a subtractive polymerase chain reaction (PCR) procedure is subsequently performed using a first primer directed against the forward adaptor sequence and a second primer directed toward the reverse adaptor sequence such that only the oligonucleotide extension product with the forward adaptor sequence at one end and the reverse adaptor sequence at the other end can be amplified and thus enriched.

Id., .0024 (17:52-60).

156. The prosecution history also supports this understanding. During prosecution of the applications that would issue as the '012 and '108 patents, the same Examiner issued obviousness rejections to the pending independent claims. EX1016.0190-96. Taking the prosecution history for the '012 patent as an example (from my review, the prosecution histories reflect the same relevant rejections and bases), the Examiner rejected the pending claim based on a combination of Jones

and Tucker. The Examiner described Jones as “a method for enriching a nucleic acid sequence of interest in a sample comprising nucleic acids” and as teaching “that nucleic acids amplified by the disclosed method may be analyzed by array hybridization or any other method known in the art to obtain sequence information about a region of interest.” *Id.*, .0190-92. There was no other step identified as “enriching” for purposes of Jones. *See id.* The Examiner also noted that it “would have been *prima facie* obvious for one of ordinary skill in the art at the time of the invention to analyze the amplification products generated by the method of Jones using one of the massively parallel sequencing methods described in Tucker.” *Id.*, .0193. The Examiner equated this analysis of “the amplification products generated by the method of Jones using one of the massively parallel sequencing methods described in Tucker” to the claimed limitation that the enriched nucleic acid sequence of interest is sequenced on a massively parallel sequencing platform. *See id.* A POSA would have understood that the Examiner considered the amplification process of Jones to result in enrichment. Additionally, while Applicants traversed the rejection, Applicants did not challenge the Examiner’s understanding that the claimed methods required enrichment through amplification. *See id.*, .0286-89. A POSA would have understood this as an agreement by Applicants that the amplification step must cause enrichment.

2. “RNA fragment”

157. This term appears in claim 21 of the '012 patent, which depends from claim 1 and further requires that the “nucleic acid fragment of the nucleic acid fragment ligated to the partial duplex adaptor in step a) is obtained by performing reverse transcription on *an RNA fragment* prior to obtaining the nucleic acid fragment ligated to the partial duplex adaptor in step a).” EX1001.0033, cl. 21. In the context of the patent, a POSA would have understood that this term is meant to have its plain and ordinary meaning (I did not see any express definition for the term), which here means an RNA polynucleotide product generated through the fragmentation process.

158. RNA is a nucleic acid. *See id.*, .0018 (6:53-54), .0025 (20:1-2). The claims and specification differentiate between a “nucleic acid” and a “nucleic acid fragment.” For example, claim 16 further requires “fragmenting a nucleic acid, thereby generating a nucleic acid fragment.” Similarly, claim 20 further requires “fragmenting a nucleic acid, thereby generating a nucleic acid fragment.” The specification describes the claimed methods as “for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising: (a) fragmenting the nucleic acids, thereby generating nucleic acid fragments.” *Id.*, .0016 (2:34-38). A POSA would understand that the patents treat nucleic acids and fragments as separate components. A POSA would also understand that the

nucleic acid fragments are generated from nucleic acids through fragmentation based on the plain language of the patent.

159. Specific to the dispute between the parties, the specification also differentiates between mRNAs and “fragments thereof.” *See id.*, .0025 (19:54-35). The ’012 patent states that the “input can be a nucleic acid” and that “[i]n one embodiment, the input can be RNA.” *Id.* (19:54, 58). The specification discloses that it can “include RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA, and fragments thereof.” *Id.* (19:60-65). A POSA would understand that an mRNA would not be an RNA fragment in the context of claim 21 but instead an RNA/nucleic acid. A POSA would also understand that the mRNA would need to be fragmented to be considered an RNA fragment.

160. This understanding is reinforced by various publications in the art. *See e.g.*, EX2061.0001 (demonstrating the cleavage of target mRNAs and the generation of “mRNA cleavage fragments”); EX2077.0001 (disclosing a protocol for RNA fragmentation and referring to the products as “RNA fragments”); EX2078.0003 (disclosing the fragmentation of polyA-containing mRNA molecules and referring to the fragmentation products as “RNA fragments”).

VII. RELEVANT PRIOR ART

A. Meyer (EX1006)

161. For all five Grounds alleged against the '012 patent and four Grounds alleged against the '108 patent, IDT relies on Meyer. Published in early 2009, Meyer reports on methods of preparing and sequencing coral larvae cDNA with the goal of assembling and annotating an entire transcriptome. *See* EX1006.0001, EX2287.¶¶ 15-20.

162. Meyer was published on May 12, 2009 in *BMC Genomics*. Meyer was a collaboration between researchers at University of Texas at Austin, Indiana University, and James Cook University. *See* EX1006.0001, EX2287.¶¶ 8-10. The study was conceived by Dr. Mikhail Matz of the University of Texas at Austin, who led the research study, oversaw the experiments and analysis, and contributed to drafting of the Meyer manuscript. *See* EX1006.0001, EX2287.¶¶ 9, 11.

163. As reported in the paper and explained by Dr. Meyer, researchers at Dr. Matz's laboratory and two other research institutions were attempting to show as proof-of-concept that it would be possible to sequence an entire coral transcriptome in one sequencing run with the 454 GS-Flx sequencing platform. EX1006.0002; EX2287.¶ 20. This was important because no reference transcriptome existed for coral species, which was a blind spot for marine biology and ecology research. EX1006.0002; EX2287.¶¶ 15-18.

164. To assist in addressing this issue, Meyer sought to develop a sequencing method for *de novo* transcriptome assembly, wherein a transcriptome is constructed from sequencing data from scratch without referencing a pre-existing reference transcriptome, using the 454 sequencing platform. EX1006.0002. As explained above, the transcriptome is the full range of mRNA expressed by an organism. *See supra* ¶ 44. These mRNA transcripts are what Meyer sought to sequence. EX1006.0002; EX2287.¶¶ 20-21.

165. Meyer applied its method to sequence larval *Acropora millepora* (“*A. millepora*”) coral samples and used the resulting sequencing data to construct a *de novo A. millepora* transcriptome. EX1006.0002. Meyer selected *A. millepora* for a proof-of-concept application of its method in part due to its status as an emerging model organism for which some pilot characterization had been conducted, but for which a comprehensive transcriptomic assessment was lacking. *Id.* Meyer hoped that this method would be applied generally to characterize the transcriptomes of uncharacterized emerging model organisms by “generat[ing] catalogs of genes and genetic markers.” *See id.*, .0001.

166. To accomplish this goal, Meyer reports on preparing a library of cDNA for sequencing, the sequencing of that library, and the subsequent analysis of the sequencing reads. Meyer discloses bespoke modifications to standard protocols to resolve known sequencing issues that arose when sequencing entire samples on the

454 sequencing platform. Meyer does not concern nor address the enrichment of targeted sequences.

1. The Specific Issues Addressed by Meyer

167. Meyer's sequencing method is specialized for use with the 454 sequencing platform. Accordingly, it sought to address several issues with the platform that reduced the quality of the resulting sequencing data and limited the viability of subsequent sequence analysis. As noted by Meyer, the 454 sequencing platform struggled to process homopolymers⁴ longer than 8 nucleotides, which resulted in errors that threw off the calibration of the device, leading to sequencing misreads. *See id.*, .0009. This issue was due to the structural design of the 454 platform, which resulted in signal broadening that prevented accurate base calling during sequencing and led to lower representation of the homopolymer regions in the resulting sequencing dataset. *See supra* ¶ 88. This issue was particularly problematic when used for transcriptome sequencing (as in Meyer). EX1006.0009, EX2287.¶ 22. This is because the polyA tails present on the 3' ends of mRNA transcripts are long homopolymers and therefore transcriptome sequencing on 454

⁴ A homopolymer is a polynucleotide sequence that contains the same repeating nucleotide, such as AAAAAAA or TTTTTTT.

platforms was structurally biased away from the accurate sequencing of the 3' ends of transcripts. *Id.*, ¶¶ 29-30.

168. Meyer also sought to address issues with adaptor concatenation during sample preparation and sequencing. EX1006.0009. Adaptor concatenation (also referred to as primer dimers) is a phenomenon where adaptors attach to each other rather than to a DNA fragment, resulting in a lower final sequencing yield due to adaptor sequestration or over-representation of attached adaptors that lack the actual DNA insert sequence. *See supra* ¶102.

169. In order to address these issues, Meyer included certain features in its sequencing method focused on specific sequencing library preparation steps. First, to address the issue with homopolymers arising from polyA tails present on the mRNA transcripts, Meyer introduced a custom (preset) cDNA primer containing a “broken-chain” region that converted one thymine to a cytosine in the middle of the poly-thymine sequence. *See* EX1006.0012-13. This was designed to facilitate incorporation of this broken chain T primer⁵ into the reverse-transcribed cDNA in order to “minimize the potential for 454 sequencing problems in this homopolymer

⁵ Highlighted in bold and underlined in the following primer sequence provided by Meyer (*see* EX1006.0012-13):

(AAGCAGTGGTATCAACGCAGAGTCGCAGTCGGTACTTTTTT**C**TTTTTTV)

stretch” by limiting the length of the resulting homopolymer sequence to less than 8 nucleotides. *Id.* Second, Meyer sought to limit the effects of adaptor concatenation by “(1) sequencing all fragments from an internal end, so that any concatenated adaptors encountered will tend to be found at the end of the read; and (2) avoiding the over-amplification that can contribute to adaptor concatenation.” *See id.*, .0009.

170. While the homopolymer optical issue and adapter concatenation were the primary concerns that motivated Meyer’s improvements over prior art techniques, there were other concerns expressed by Meyer that dictated specific design choices. EX2287.¶ 22. Because Meyer sought to sequence just mRNA transcripts, Meyer needed to isolate the mRNA from total RNA. EX1006.0009; EX2287.¶ 71. This was accomplished through the SMART protocol, which allowed for the selection of mRNA over total RNA. EX1006.0012-13; EX2287.¶ 71. Because Meyer wanted to sequence the entire transcriptome, Meyer needed to overcome an issue in reverse transcription where the reverse transcriptase did not always complete copying of the mRNA transcript. EX1006.0002; EX2287.¶¶ 24-25. Reverse transcription begins at the 3’ end, so any failure to copy the entire transcript would lead to an under-representation of the 5’ end and an over-representation of the 3’ end. EX2287.¶¶ 24-25. To address this issue, Meyer relied again on the SMART protocol, which was designed to enrich for full-length strands given its use of the CAP primers. EX1006.0012-13; EX2287.¶ 66. Finally, because

PCR amplification tends to bias in favor of shorter strands, Meyer needed to ensure that cDNA strand sizes were normalized before exponential amplification. EX1006.0012-13; EX2287.¶ 26. This was accomplished through Evrogen's Trimmer protocol. EX1006.0012-13; EX2287.¶¶ 23-26; EX2081.0003. As discussed further below, all of these processes occurred before the cDNA strands were fragmented and before any partial duplex adapters were ligated to the cDNA fragments.

2. cDNA Library Preparation

171. The Meyer library preparation method used to sequence the *A. millepora* transcriptome using the 454 sequencing platform, which for clarity I refer

to as the “general method,” is disclosed in Figure 1, reproduced below. See EX1006.0002-03; *see also id.*, .0012-14.

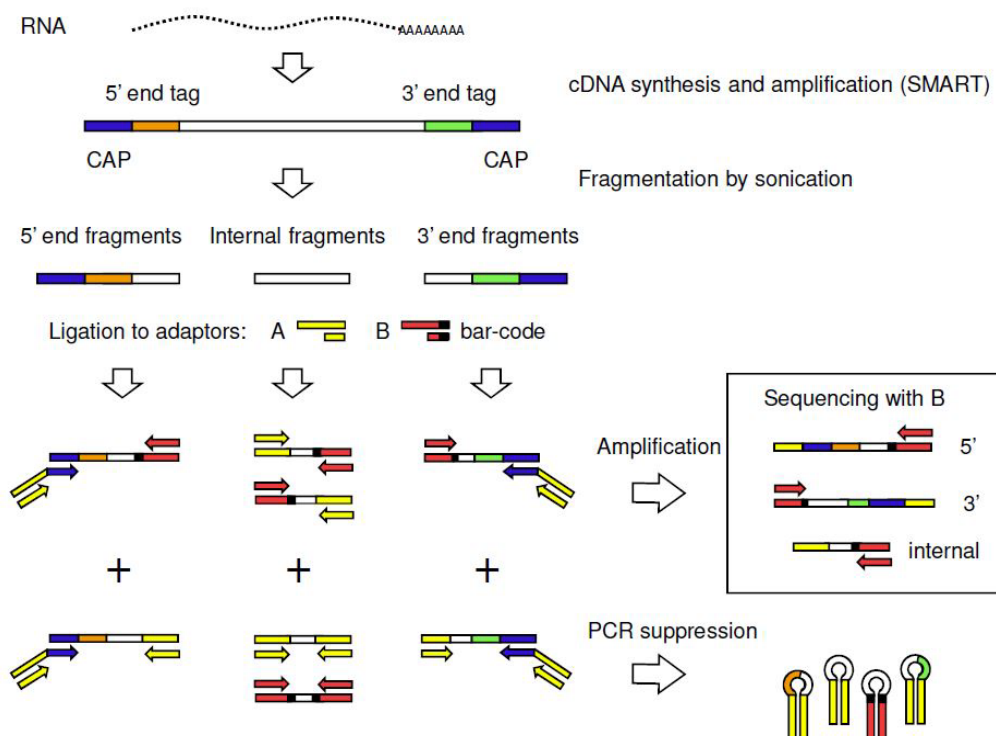
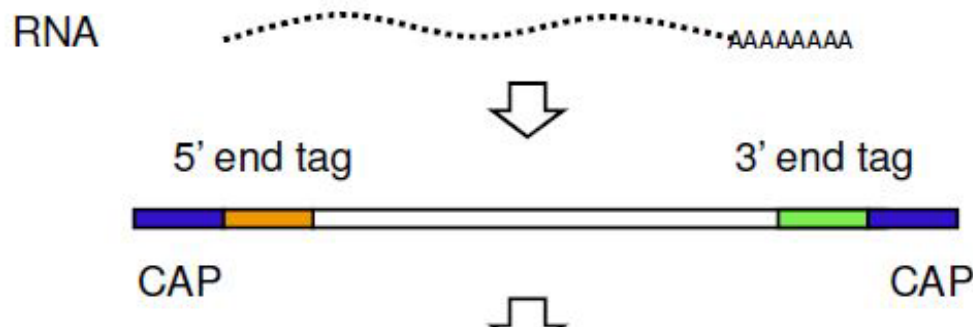


Figure 1
Diagram of cDNA synthesis and 454 library preparation procedures. Three fragment types are produced by sonication: 5', internal, and 3'. Ligation with the partially double-stranded adaptors A and B produces, for each fragment type, certain adaptor configurations that will be amplified (above the '+'), and others that will be suppressed (below the '+') during the subsequent amplification. The procedure preferentially amplifies constructs that are appropriate for 454 sequencing (shown inside box).

EX1006.0003.

172. Meyer applied the general method as follows. RNA samples, in this case larval *A. millepora* samples exposed to several different treatments designed to model stress environments, were first generated by RNA extraction to isolate the RNA from the tissue sample. *See id.*, .0012. Meyer purified these RNA extracts to

minimize unwanted samples. *See id.*, .0012-13. After purification, Meyer reverse transcribed the RNA into cDNA using Clontech's SMART cDNA synthesis protocol.



173. The cDNA synthesis process is critical to the library preparations of Meyer for several reasons.

174. First, Meyer uses this process to address the homopolymer optical issue of the 454 sequencer. The SMART protocol requires the use of two cap primers, which target the termini of mRNA transcripts. One of the cap primers is directed to the polyA tail of mRNA. Meyer customized this primer by changing one nucleotide in the primer from a “T” to a “C.” By doing so, Meyer was able to use this primer to truncate the polyA tail, which can be hundreds of nucleotides in length, and to disrupt the corresponding polyT chain in the synthesized cDNA. *See id.*, .0012-13. Meyer hypothesized that by disrupting the polyT chain, breaking the homopolymer into smaller homopolymers, they could fix the optical issue of the 454 sequencer. *Id.*

175. Second, Meyer uses the cap primers to promote read directionality. These cap primers are designed so that sequences cannot be ligated to their ends. *Id.*,

.0019; EX2287.¶¶ 37-40. This means that once a transcript is “capped,” no ligation will occur at the transcript’s termini. When Meyer later fragments the transcripts, these caps allow for the identification of the end fragments and also for Meyer to control for which adapter sequence is incorporated at the termini (more on this below).

176. Third, Meyer uses the SMART protocol to enrich for mRNA. Using the SMART protocol with a cap primer specific to the polyA tail allowed for Meyer to target mRNA over other types of RNA in the sample, such as ribosomal RNA. EX2287.¶¶ 24-25. Meyer amplified the capped strands as part of cDNA synthesis (and before fragmentation) to enrich for mRNA in the sample. *Id.*

177. Fourth, Meyer uses the SMART protocol in conjunction with the Evrogen Trimmer Kit to also enrich for longer, full-length strands of mRNA. *See* EX1006.0012-13. Unaided amplification often biases towards the amplification of shorter polynucleotides over longer ones. EX2287.¶ 26; EX2081.0009. There are several reasons for this, but one common reason is that longer polynucleotides are more likely to suffer from concatenation issues, such as by the ends ligating to each other and forming a circularized structure. The SMART protocol addresses this PCR bias through having the cap primers target the 5’ and 3’ termini of mRNA transcripts. Only transcripts with both ends (i.e. full length) should be capped. EX1006.0012-13; EX2287.¶ 26; EX1034.0004-05. The Evrogen Trimmer Kit allows for

normalization of these transcripts to promote full-length strands. EX1006.0013; EX2287.¶ 37; EX2081.0003. The SMART protocol then uses primers specific to the cap primer sequences to amplify those capped transcripts, thus enriching for the capped transcripts over other RNA in the solution. EX1006.0013; EX2287.¶ 35; EX1034.0004-05.

178. By the end of this process, Meyer discloses a purified library of normalized cDNA strands that correspond to full-length mRNA transcripts. *See* EX1006.0013. These cDNA strands are capped at the 5' and 3' ends, with the 3' end containing a modified cap sequence that incorporates the broken-T chain. As explained above, by using the SMART protocol, Meyer is able to address the under-representation issues of the 5' ends that occur as a result of reverse transcription. This enrichment for full-length strands also addresses any over-representation of the 3' ends that can occur from short, premature copying during reverse transcription. EX2287.¶¶ 24-25.

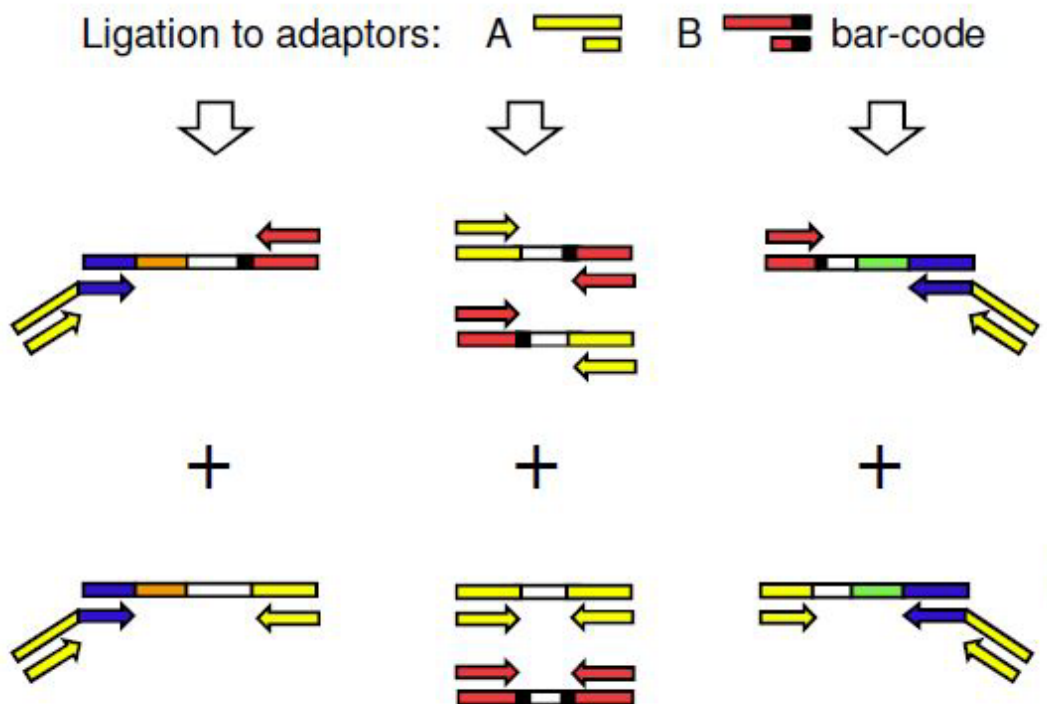
179. After cDNA synthesis and amplification, the cDNA strands were next fragmented. Meyer discloses fragmentation by a sonication procedure designed to produce random fragments approximately 300–400 nucleotides in length. EX1006.0013. The fragmentation process produced a mix of 5' fragments that contained the cap region, internal fragments, and 3' fragments that contained the cap region and the broken-chain region, pictured below.

5' end fragments Internal fragments 3' end fragments



Id., .0003.

180. After fragmentation, Meyer discloses a ligation step. Two partial-duplex adaptors containing discrete primer sequences, “A” and “B,” were then ligated randomly to the fragments. *See id.*, .0013. The cap regions were designed such that ligation could not occur at the end containing the cap region, resulting in only one adaptor ligating to the end fragments rather than two. *Id.* The adapters are added in excess to promote adapter ligation to all fragment ends no that did not contain a cap. *Id.* This is reflected in Figure 1.



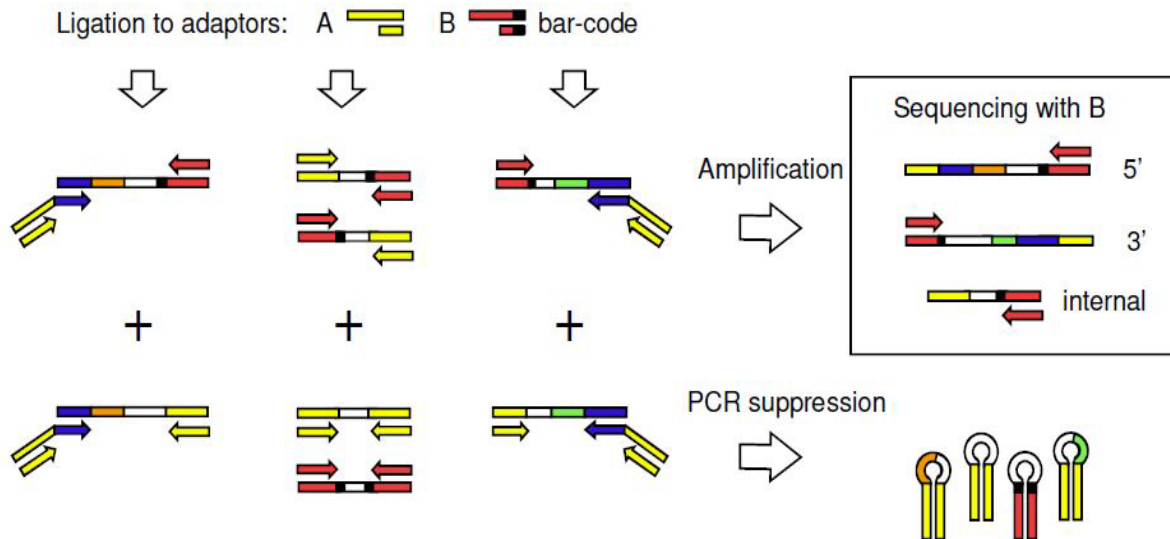
Id., .0003. As reflected in the figure, there are eight possible configurations. For the 5' end fragments, the adapter-ligated fragments contain the cap primer at the 5' end and either the Primer A or Primer B sequence at the 3' end. For the internal fragments, the adapter-ligated fragments contain either the primer sequence combinations (from 5' to 3') A/B, B/A, A/A, or B/B. A POSA would understand that the primer binding sequences are present on both strands due to polymerases filling in the complement strand as denaturation is about to occur. *See id.*, .0013. For the 3' end fragments, the adapter-ligated fragments contain either the Primer A or Primer B sequence at the 5' end and the cap primer with the broken-T chain at the 3' end.

181. The ligation design implemented in Meyer was also designed to ensure that subsequent sequencing could be conducted using a single primer (primer B) while maintaining appropriate read orientation. *See id.* This is achieved by utilizing a PCR suppression approach during the amplification step that follows adapter ligation. *Id.* The adapters were designed such that fragments with the same adapters on both ends would interact with each other. *Id.* This interaction prevented primer binding and therefore subsequent amplification.

182. After adaptor ligation, Meyer discloses the amplification of the cDNA fragments. In the first step of amplification, Meyer anneals specific primers to the cDNA fragments. To account for the capped fragments (both end fragments), Meyer utilized a mix of three primers, “A primer,” “B primer,” and “A +CAP primer” that

annealed to the A adaptor, B adaptor, and cap regions respectively to generate extension products of the fragments. *Id.* The A +CAP primer included the A primer sequence at its 5' end, which allowed for extension products from the A +CAP primer to contain the A primer sequence as well. *Id.* Because all three primers were present in the solution, extension products are generated for both end fragments and internal fragments.

183. The extension products are then amplified using the A and B primers. *Id.* Amplification occurred in an unbiased manner, as all non-PCR-suppressed fragments were amplified by the A and B primers. *Id.* The amplified and suppressed fragments, respectively, are shown below.



EX1006.0003. As can be seen, all fragments above the three “+” symbols undergo amplification. Meyer discloses that this amplification is exponential and that the “[a]mplification was carried out for 17 cycles” before “repeating the amplification

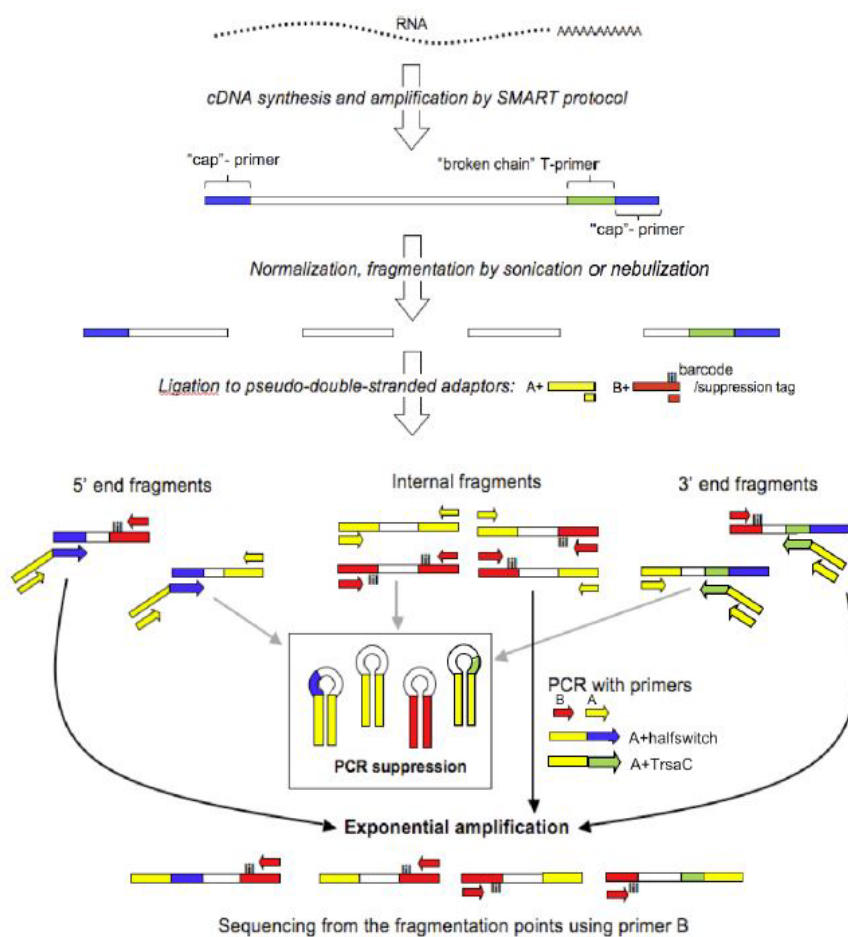
using a 5'-biotin-labeled 'A' primer," which is typically used to bind the amplicons to beads for 454 sequencing. *See id.*, .0013. The amplification products are depicted in the right box of the figure. That box shows that the amplification products include 5' end fragments, internal fragments, and 3' end fragments. The box also shows that sequencing starts at the Primer B sequence.

184. For the fragments below the three "+" symbols, Meyer discloses that the fragments are PCR suppressed. This is because the fragments contain identical flanking adaptor sequences. As can be seen from the figure, these fragments bend into a hairpin loop because of the matching sequences, which suppresses amplification.

185. I understand that IDT relies on the modified method depicted in Additional File 4 rather than the general method. *See* Paper 1 (IPR2025-00015), 28; Paper 1 (IPR2025-00016), 27. I also understand that Dr. Sims reaffirmed his reliance on the modified method at his deposition. EX2037.0021 (221:20-225:10). As disclosed in the publication and explained by Dr. Meyer, Meyer relied on the cDNA preparation method disclosed in Figure 1 to prepare the coral samples for sequencing. EX1006.0003; EX2287.¶¶ 63-64. While Additional File 4 does detail specifics about the protocol depicted in Figure 1, it also "includes minor modifications . . . recently developed to reduce the number of manipulations required, improve reproducibility, and to further reduce adaptor concatenation." *See*

EX1006.0014. The modified method was not used to produce any of the data disclosed in Meyer. *Id.*, .0019 (“the version of the protocol shown here includes minor modifications from that used for preparing coral larval cDNA.”).

186. Meyer includes a schematic of this method in Additional File 4, which is reproduced below:

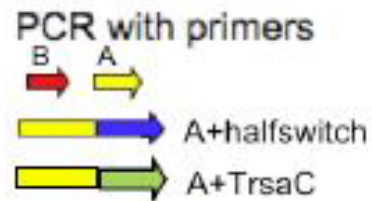


Id., .0020.

187. As shown above, the modified method largely follows the same scheme as the general method illustrated in Figure 1. Specifically, the cDNA synthesis and amplification by SMART protocol, normalization, fragmentation, and adaptor ligation steps are the same in both methods. *See id.*, .0020-24. One minor difference to note is that Meyer discloses in the modified method that fragmentation can occur either by sonication or nebulization. Sonication uses ultrasonic waves to fragment DNA through resonance vibration. *See* EX2070.0001. Nebulization uses compressed air to force DNA through a small hole, which fragments the DNA in the process. *Id.* Studies have shown that both methods perform roughly about the same with “only minor differences.” *Id.*

188. The methods disclosed in Meyer vary substantively following the ligation step, where the modified method extends the ligated fragments using 4 primers rather than the 3 primers used in the general method. *See* EX1006.019-20. Two of the primers, the A and B primers, are the same from the general method. *Id.* One primer, the “A+halfswitch” appears to be the same primer that was used in the general method to incorporate the Primer A sequence onto the termini of the end fragments. *Id.*, .0013, 0019-20. This primer is identified differently in the figure depicting the modified method, but I have not discerned any substantive difference between it and the primer targeting the cap primer sequences in Figure 1. The final primer, the “A+TRSAC” primer appears to be new to the modified method. While

this primer serves the same function as the “A+halfswitch” primer and includes the Primer A sequence at its tail, the “A+TRSAC” primer contains sequence complementary to the broken-T chain rather than the cap primer. The four primers are depicted below.

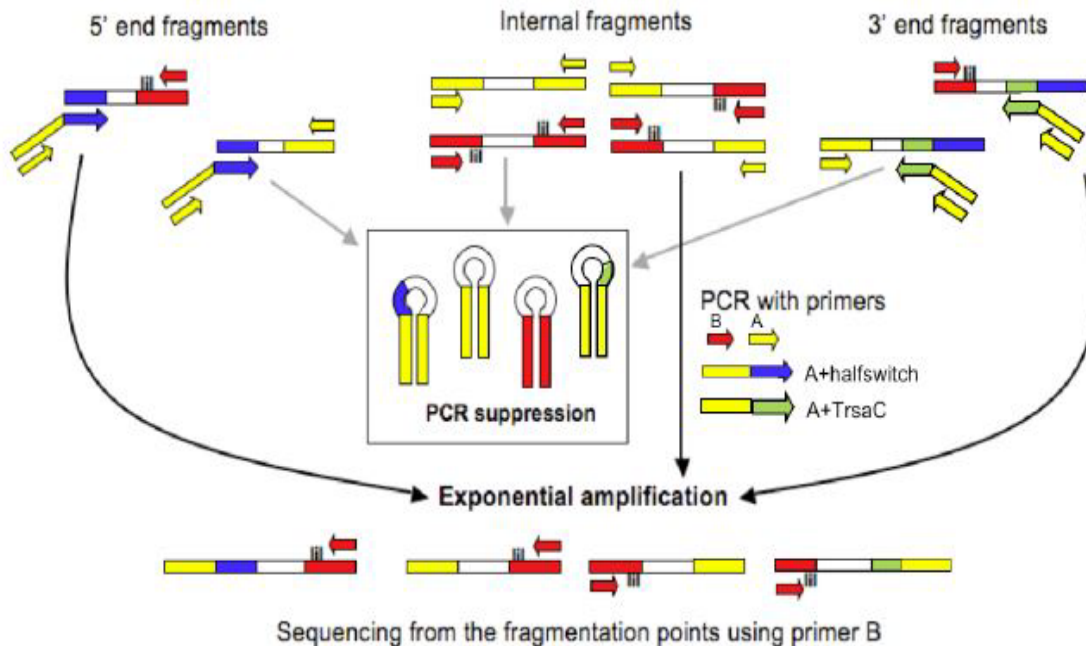


Id., .0020.

189. By using the “A+TRSAC” primer, the modified method generates a different extension product for the 3’ end fragments than the general method. The 3’ end fragment extension products generated by the modified method do not contain the cap region, as the cap region is located downstream of the broken-chain region bound by the A +TRSAC primer. *Id.* Otherwise, the extension products of the general and modified methods are the same. *See* EX2287.¶ 49.

190. Despite this variation in primers, and therefore the extension product composition, the modified method retains the general amplification workflow of the general method. *See* EX1006.0025; EX2287.¶¶ 47-49. All extension products are amplified in an unbiased manner using the standard A and B primers, where all fragments are amplified or suppressed depending on the primer sequences located at

both ends of the fragments. EX1006.0020, 25; EX2287.¶¶ 48-50. This process is reflected in the below diagram.



EX1006.0020. As with the general method, the modified method relies on PCR suppression to prevent the exponential amplification of fragments containing identical adaptor sequences on both ends. EX1006.0003, 0020; EX2287.¶¶ 45, 48. The result of the exponential amplification is a set of amplicons that correspond to the 5' end, all internal fragments, and the 3' end. EX1006.0020; EX2287.¶ 71. Except for the 3' end amplicons not containing cap primer sequence, the amplicons are identical to those of the general method. EX1006.0003, 0020.

3. Sequencing the cDNA Library

191. After amplification of the cDNA fragments by the general method in Figure 1, Meyer discloses that the library undergoes purification and titration before

sequencing. *Id.*, .0013. Meyer explains that the purification process includes size-selection and biotin-streptavidin purification, which ensure that amplicons less than a certain length are not sequenced and that only amplicons with the correct adaptor A/B constructs are sequenced. *Id.*, .0014. Meyer explains that the biotin-streptavidin purification was not necessary for their method because the “library was essentially free of incorrect adaptor configurations.” *Id.* Meyer discloses that the “sample was titrated to optimize yield and sequence quality.” *Id.* This titration process incorporated the emPCR process used for 454 sequencing. After this process is completed, Meyer sequenced the prepared amplicons “using the 454 GS-Flx instrument according to the manufacturers’ instructions.” *Id.*

4. Data Analysis

192. Meyer reports that their method yielded 628,649 raw sequencing reads from a single sequencing run of the 454 GS-Flx platform. EX1006.0002; EX2287.¶ 51. The raw reads correspond to the number of sequencing reads that were received back from the 454 sequencer before any further quality controls were performed. After adapter trimming and size selection, Meyer assembles a portion of the raw reads into 44,444 contigs. EX1006.0003. A contig refers to the joining of multiple, overlapping sequencing reads into one piece. EX2287.¶ 52.

193. Meyer reports that after assembly of the contigs, they were left with 62,657 singletons. EX1006.0003. A singleton is a single sequencing read, which

often cannot be joined to form a contig. EX2287.¶ 53. Meyer reports that their analysis of the singletons showed that “a substantial fraction” were in fact “artifacts of cDNA synthesis, normalization, or sequencing.” EX1006.0010. Singletons comprised 10% of their sequencing reads. *Id.*, .0010. Meyer explained that even though 10% of their raw reads were singletons, it was an improvement over other recent 454 transcriptome assemblies where the number of singletons ranged from 13 to 60%. *Id.*

194. Meyer compared the contigs and singletons to known protein sequences. EX1006.0005; EX2287.¶ 54. Because a few thousand singletons returned matches to known proteins, they combined the contig and singleton matches to construct clusters of nucleic acid sequences that allowed for them to reconstruct the coral transcriptome. Where assembled sequences matched genes in public databases, they annotated the genes to demonstrate coverage. EX1006.0006.

195. Meyer evaluated the completeness and depth of coverage of the generated transcriptome library in several ways. First, they confirmed that all genes in major metabolic pathways were identified in our library. *Id.* The presence of these genes in their library suggested that the “annotated sequences account[ed] for nearly the complete coral larval transcriptome.” *Id.* Second, they confirmed the presence of regulatory genes that had been identified in the most closely related organism for which an assembled draft of the genome sequence was available. They

were able to confirm the presence of 23 out of 24 sets of genes. *Id.*, .0007. They did the same for transcription factors and identified matches for 18 out of 20 families of sequences. *Id.* These comparisons indicated to them that “the collection of sequences established in this study provides a reasonably complete description of the coral larval transcriptome.” *Id.* Finally, as mentioned above, they compared the sequencing read data obtained in their study to that from other studies. According to Meyer, these comparisons highlighted “the potential for our method to improve sequencing depth and efficiency in future transcriptome sequencing efforts.” *Id.*, .0010.

196. To test how effective the method was in resolving the homopolymer issue, they examined how many raw reads contained a single copy of the “broken-T” chain sequence. EX1006.0009; EX2287.¶ 55. Of the 628,649 raw sequencing reads (EX1006.0002), 98,276 raw reads contained a single copy of the “broken-T” chain. That is approximately 15.6% of total raw reads (Meyer reports 16%). EX1006.0009. They compared that percentage to an “expectation” of approximately 10%, which according to Meyer, was “based on average read length (232 bp) and assuming an average transcript size of 2,200 bp.” *Id.* Meyer does not report from where the 2,200 bp average came *Id.*

197. Finally, they also evaluated the number of single nucleotide polymorphisms (“SNPs”) that were present in the sequencing data. EX1006.0012,

EX2287.¶ 59. Using a newly developed software program (QualitySNP), they identified around 30,000 potential SNPs. EX1006.0009. To confirm the accuracy of the predicted SNPs, they selectively enriched for and sequenced 20 SNPs. *Id.*; EX2287.¶ 60. They used target-specific primer pairs to target and enrich for each SNP. *Id.* These primer pairs were reported in Additional File 3, which I understand was not included in Exhibit 1006. I also understand that Additional File 3 has now been submitted as Exhibit 2079. I have reviewed this exhibit and confirm that Additional File 3 reported target-specific primer pairs for each SNP. EX2079. After they selectively enriched for the 20 SNPs, they confirmed through Sanger sequencing that 70% of the predicted SNPs were accurate. EX1006.0009; EX2079.0001.

B. Siebert (EX1007)

198. Published in 1995, Siebert claims to disclose “[a]n improved PCR method for walking in uncloned genomic DNA.” EX1007.0001. As explained above, this publication is one of the earliest studies that became referred to as “genome walking.” *See supra*, ¶ 62. The primary goal of genome walking was to map unknown regions of the genome. *Id.* Many genome walking methods relied on the targeting of known regions of the genome, which could then be used to generate products that were analyzed by gel electrophoresis. *Id.* Siebert was no exception.

199. Siebert explains that there were three “types” of PCR-based methods that could be used “for walking from a known region to an unknown region in cloned or uncloned genomic DNA”: inverse PCR, randomly primed PCR, and adaptor ligation PCR. EX1007.0001. But as Siebert explains, “these methods have not been generally applied to walking in uncloned genomic DNA because they are either complicated or inefficient.” *Id.* While Siebert reports that recent improvements have helped, the improvements have only led to limited genome walks. *Id.*

200. Seeking to improve on PCR-based methods, Siebert discloses that they investigated using “unpredictably primed PCR” with “vectorette PCR.” *Id.* A POSA would have understood “unpredictably primed PCR” to refer to a PCR method wherein a genetic sample is characterized by randomly annealing primers targeting common repetitive sequence elements. A POSA would have understood “vectorette PCR” to refer to a method of PCR where a special adapter, known as a vectorette, is used in conjunction with a single target-specific primer to amplify specific segments of DNA. EX2044.0001. The vectorette is designed with a mismatched sequence. *Id.* When the target-specific primer is extended to the end of the vectorette, a binding site is created that allows for subsequence amplification. *Id.*

201. Siebert reports that both UP-PCR and vectorette PCR “generated multiple PCR products even after nested PCR was performed.” EX1007.0001. According to Siebert, the multiple PCR products were undesirable because they

“would complicate further characterization without Southern blot hybridization or extensive cloning.” *Id.* Southern blot hybridization relies on hybridization probes to detect for specific DNA sequences.

202. Based on these experiences, Siebert discloses that they “have improved upon the adaptor ligation method by combining ‘vectorette PCR’ with a newly developed method termed ‘suppression PCR.’” EX1007.0001. Siebert discloses that they used specific enzymes to digest and fragment genomic DNA. *Id.* Special vectorette adaptors were then ligated onto the ends of each fragment. *Id.* These vectorette adaptors contained a specific adaptor primer. *Id.* If any PCR products were generated containing these primer sequences at both ends, the products would bend to form “panhandle” structures, which would “suppress exponential amplification.” *Id.*

203. Siebert discloses that PCR suppression did not occur for extension products that “contain the adaptor sequence on one end.” *Id.* This occurred only where the “gene-specific primer extends a DNA strand through the adaptor,” which like traditional vectorette PCR, would result in a binding site that could be used for exponential amplification. *See id.*

204. After generating PCR amplification products, Siebert discloses that the products were analyzed by gel electrophoresis. *Id.* This allowed for Siebert to determine the size of PCR products and to map those sizes against restriction enzyme

patterns to make predictions about the surrounding genome. *See id.* Siebert attributes “the high specificity of the method” to “the suppression effect of the adaptor/adaptor primer combination” but explained that vectorette blocking sometimes “yielded better results.” *Id.* Based on these results, Siebert explains that the “improved DNA walking method described here will be valuable for rapidly finding promoters and regulatory elements from sequences obtained from cloned cDNAs, to determine the exon-intron boundaries of genes, and to walk upstream or downstream from sequence tagged sites generated from human genome studies.” *Id.*

C. Caruccio (EX1008)

205. Published in 2011, Caruccio highlights Epicentre Biotechnologies’ Nextera technology. Nicholas Caruccio, the author of the reference, was the Director of Product Development for Epicentre at the time. EX2056.0001. As the publication explains, Nextera technology uses hyperactive transposase enzymes (i.e., Transposome) to fragment and tag DNA in a “near-random” reaction. EX1008.0001. A POSA would have understood that this technology facilitated fragmentation “via *random*, staggered double-strand DNA breaks in the target DNA” sample. EX2056.0001 (emphasis added).

206. Caruccio discloses that “[u]sing in vitro transposition to simultaneously fragment and tag DNA in a single-tube reaction is a significant improvement upon current procedures, which generally [have] consist[ed] of distinct DNA

fragmentation, end-polishing, and adaptor-ligation steps.” EX1008.0005. Caruccio criticizes those procedures as “often involv[ing] significant sample loss and hands-on time.” *Id.*, at .0001. Caruccio also notes that those procedures “suffer from limited sample throughput.” *Id.*, at .0002. Caruccio teaches that these issues can be reduced by using Nextera technology, which “can be exploited to perform all of these steps in a single, 5-min reaction.” *Id.* Caruccio refers to this process of random fragmentation and tagging as “tagmentation.” *Id.*

207. Caruccio discloses general library preparation methods using Nextera for the 454 sequencer (EX1008.0005-09) and Illumina (EX1008.0009-13). None of the disclosed methods use target-specific enrichment.

D. Bronner (EX1009)

208. Published in 2009, Bronner discloses “improved protocols for Illumina sequencing.” EX1009.0001. Bronner teaches a general PCR amplification scheme in its “Basic Protocol 4” where adapters are ligated to all fragments and then amplified. *Id.*, .0015; *see also id.*, .0012 (showing that adapter ligation occurs without regard for the target sequence). As an “alternative protocol,” Bronner teaches that a pair of gene-specific primers with tailed universal primers could be

used to amplify specific fragments. *Id.*, .0017-18. Bronner fails to disclose any library preparation methods directed to the use of a single target-specific primer.

209. Bronner also teaches a “PCR-free library method.” *Id.* That method does not employ any target specific primers or an amplification scheme to enrich for target sequences.

VIII. MEYER FAILS TO DISCLOSE THE CLAIMED INVENTIONS

210. I have compared the claims of the '012 and '108 patents with Meyer and determined that Meyer fails to disclose several of the claim limitations required by both patents.

211. As explained below, for claim 1 of the '012 patent, Meyer fails to at least disclose “amplifying the one or more oligonucleotide extension products . . . to enrich for the nucleic acid sequence of interest.” For claim 1 of the '108 patent, Meyer fails to at least disclose “amplifying the one or more oligonucleotide extension products . . . thereby enriching the nucleic acid sequence of interest.”

212. Because Meyer fails to disclose these limitations, it is my understanding that Meyer does not anticipate the independent claims of the '012 and '108 patents. It is also my understanding that Meyer therefore cannot and does not anticipate any of the dependent claims of the '012 and '108 patents.

A. “amplifying the one or more oligonucleotide extension products . . . to enrich for the nucleic acid sequence of interest” ('012 patent) / “amplifying the one or more oligonucleotide extension

**products . . . thereby enriching the nucleic acid sequence of interest”
(’108 patent)**

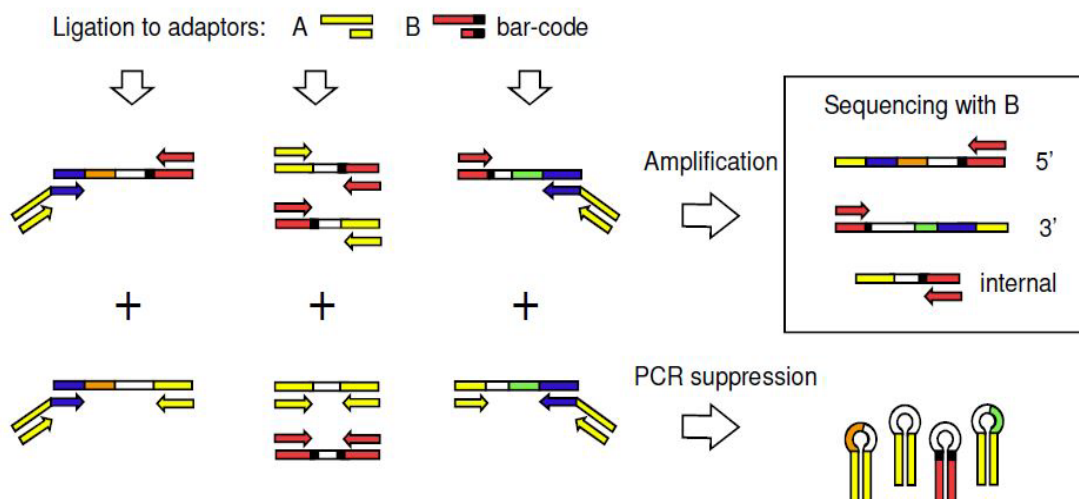
213. Meyer does not disclose the claimed enrichment of a nucleic acid sequence of interest.

**1. The Disclosed Protocols Do Not Selectively Amplify for
cDNA Fragments with Sequences of Interest**

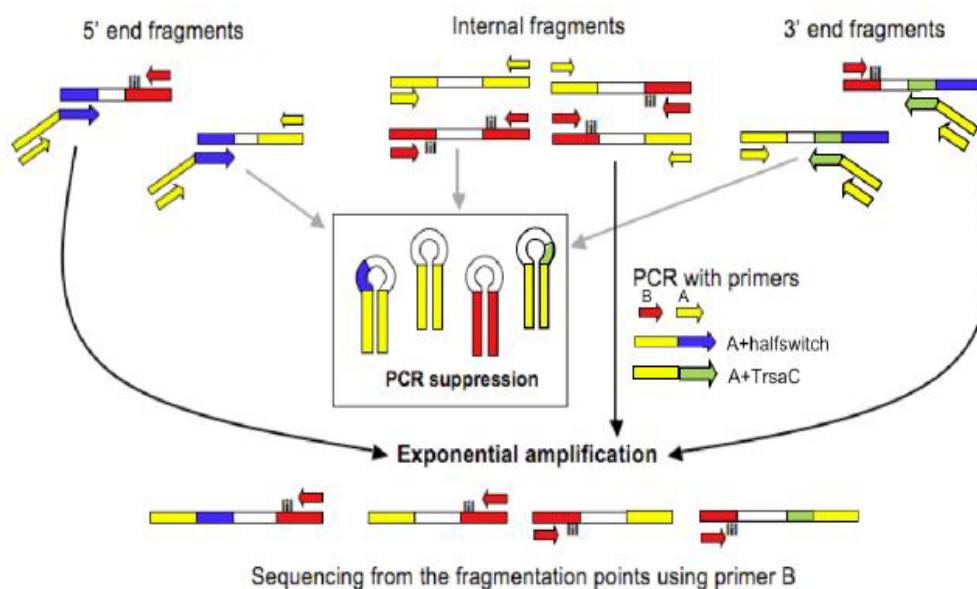
214. The ’012 and ’108 patents require that the amplification step of claim 1 result in enrichment. *See supra* ¶¶ 149-56. For both patents, that enrichment occurs as a result of the primer binding sites that are unique to the oligonucleotide extension products generated in the previous step. *Id.* The patents require that the oligonucleotide extension products contain the sequence complementary to the sequence of interest, sequence complementary to the first adapter sequence, and the second adapter sequence. *Id.* Because the oligonucleotide extension products are formed by extending oligonucleotides specific to the target sequence—and no other sequence—all oligonucleotide extension products must contain sequence complementary to the sequence of interest. *Id.* This restriction means that only the extension product sequence and its complement undergo exponential amplification. *Id.*

215. Neither method disclosed in Meyer (the general method and the modified method) disclose this amplification process. In both methods, Meyer discloses “exclusive amplification of the fragments flanked by both A and B

primers.” EX1006.0020. That “exclusive amplification” occurs for fragments representing the entire length of the cDNA strands, which is reflected in both methods:



Id., .0003 (Figure 1 depicting the general method).



Id., .0020 (Additional File 4 depicting the modified method).

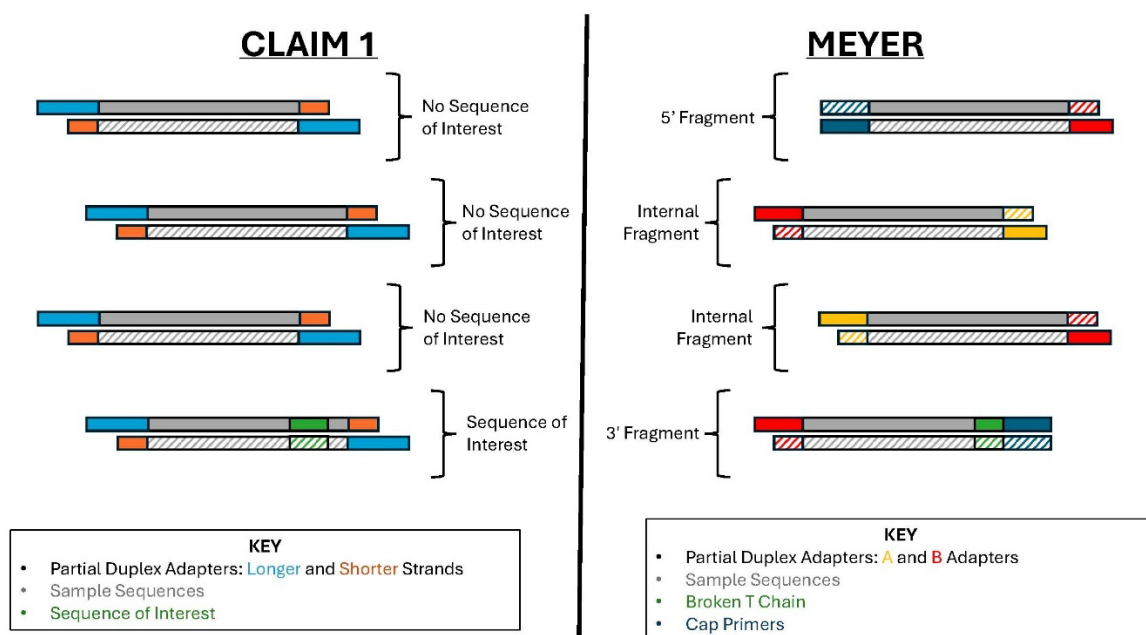
216. Starting with the general method, Figure 1 shows four different fragment configurations above the three “+” symbols. Those configurations represent the entire length of the full-length cDNA strands. After extension, all of those fragments contain the Primer A sequence on one end and the Primer B sequence on the other end. Meyer discloses that because these fragments contain both sequences, they are exponentially amplified. The box on the right of the figure shows the resulting amplicons corresponding to the 5’ end, internal, and 3’ end. Similarly, the figure in Additional File 4 shows four configurations that undergo exponential amplification. Those four configurations are the same four from the general method, except that the 3’ end configuration will no longer contain the cap primer sequence after extension.

217. Meyer also discloses four configurations of fragments in each method that do not undergo exponential amplification. Those four configurations all have identical flanking primer sequences and are thus PCR suppressed. It is important to note that both 5’ and 3’ end fragments are suppressed along with internal fragments. Meyer does not disclose any methods for biasing PCR suppression, meaning that approximately an even proportion of each fragment type will be suppressed. Dr. Meyer confirms this lack of PCR biasing. EX2287.¶ 45.

218. But even if PCR suppression resulted in a change in the relative proportions of the fragments (it does not), the amplification process disclosed in

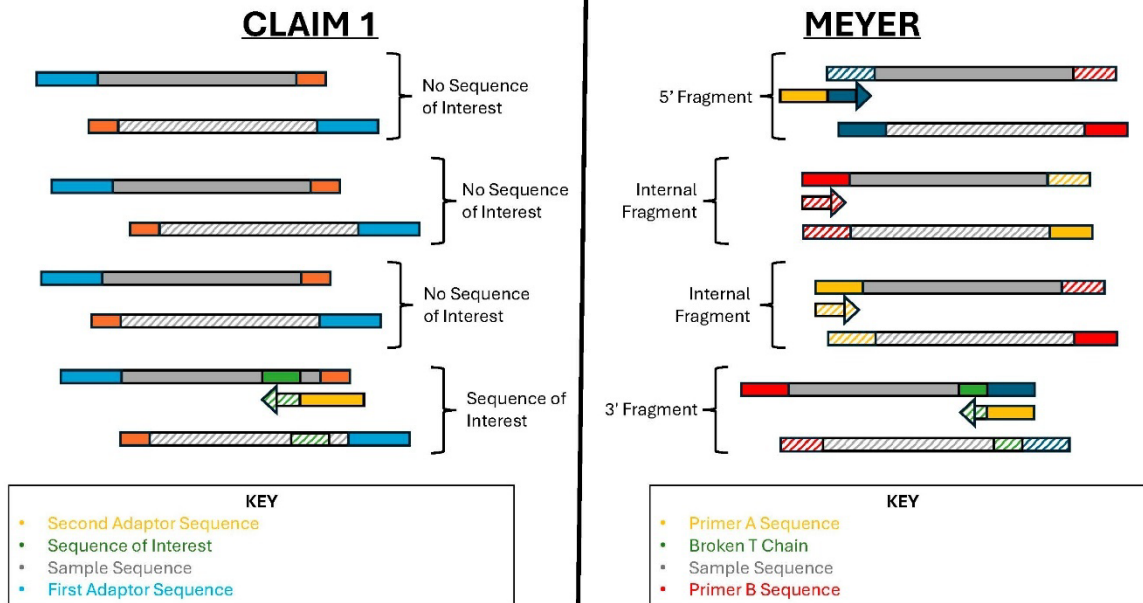
Meyer is not the same as the claimed amplification process. To illustrate why that is the case, I have prepared a series of demonstratives comparing the claimed methods to Meyer.

219. In the first step of the process, both claim 1 of the '012 patent and Meyer obtain fragments with at least one partial duplex adaptor ligated to one end of the fragment. While I disagree with IDT's claim that the broken-T chain is a "sequence of interest"—given that Meyer does not disclose a desire to enrich for it through the amplification of the cDNA fragments, I compare the broken-T chain of Meyer to the target sequence of interest in the claimed methods. All sequences are shown in solid colors, complementary sequences (excluding the shorted strand of the claimed duplex adapter) are denoted in corresponding colored patterns.



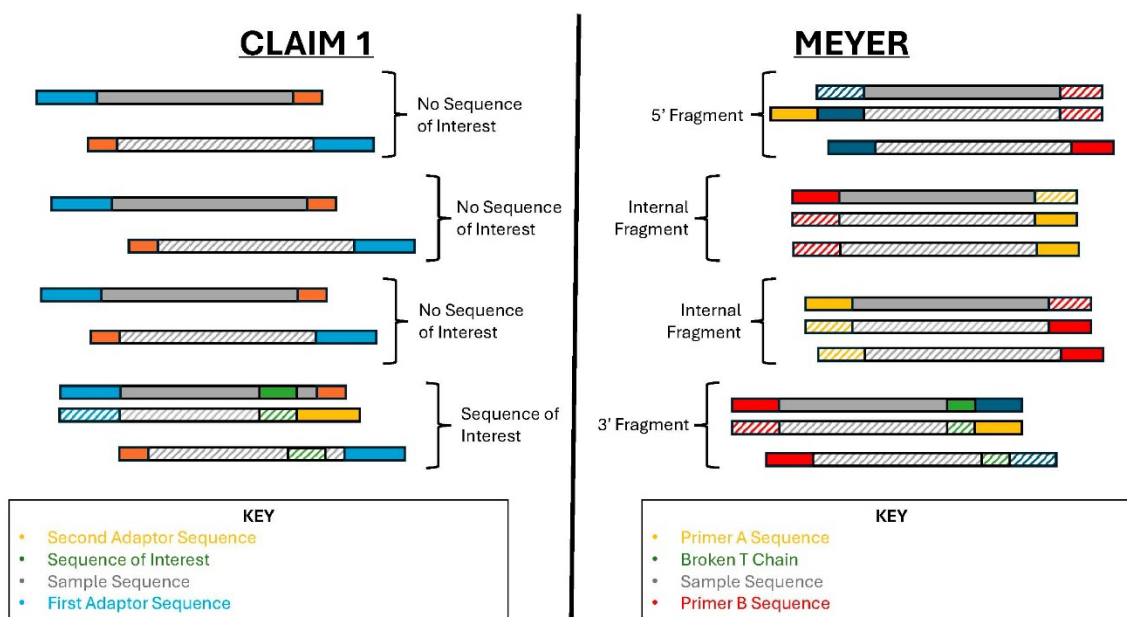
220. As can be seen, all fragments of the claimed methods are ligated to at least one partial duplex adaptor, including the fragment containing the sequence of interest (shown in green on the left panel). The partial duplex adaptors of the claimed method contain a first adapter sequence (shown in blue on the left panel). For Meyer, the end fragments can only have one partial duplex adapter ligated to their interior ends due to the cap primers at their termini. The internal fragments have partial duplex adapters at each end, with one end containing a first adapter sequence (the Primer A sequence, shown in yellow on the right panel) and the other end containing a second adapter sequence (the Primer B sequence, shown in red on the right panel). The PCR-suppressed end fragments are ligated to adaptors with the Primer A sequence (not shown). The PCR-suppressed internal fragments have matching adaptor sequences (either A/A or B/B, not shown).

221. In the next step of the process, the claimed methods target the sequence of interest with an oligonucleotide that is complementary to the sequence of interest (shown as the patterned green arrow on the left panel). This oligonucleotide contains a second adaptor sequence at its tail.



222. IDT relies on the modified method and claims that the oligonucleotide that is complementary to the broken-T chain is the equivalent component in Meyer (shown as the patterned green arrow on the right panel). That oligonucleotide, however, contains the same adaptor sequence already present on the internal fragments (the Primer A sequence, shown as the yellow bar following the green arrow on the right panel). In the modified method, Meyer also targets the cap primer of the 5' end fragment with an oligonucleotide that also contains the same Primer A sequence at its tail (shown as the blue arrow and yellow bar on the right panel). For the general method, Meyer just targets the cap primer sequences at the 5' and 3' ends (not shown in the above figure). There is no targeting of the broken-T chain.

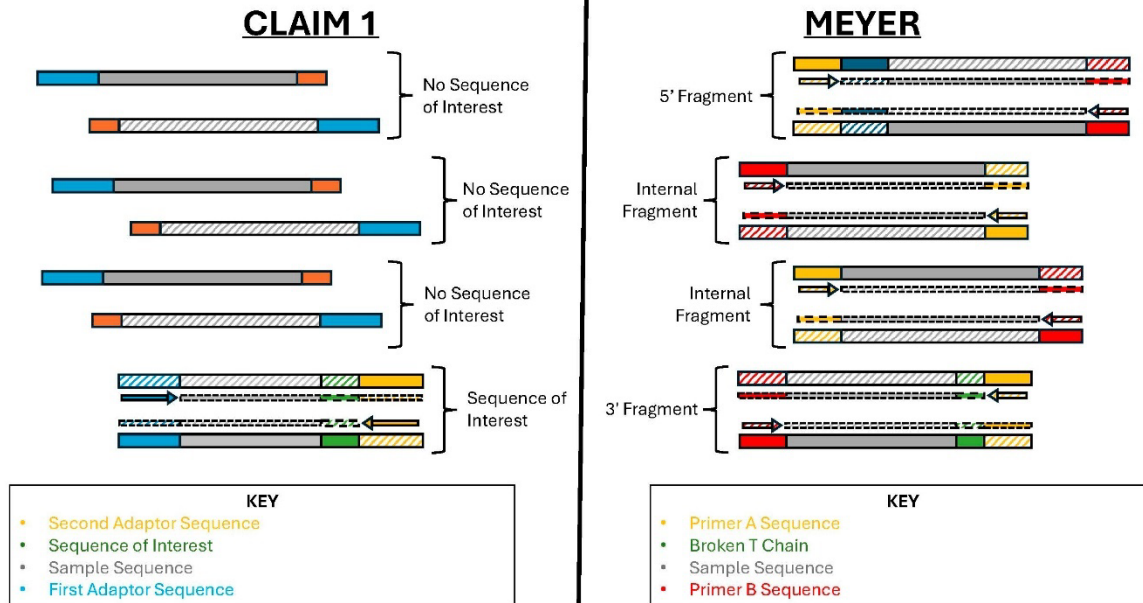
223. The distinction between Meyer and the claimed methods is further reflected in the extension step. In the claimed methods, extension only occurs for the oligonucleotide complementary to the sequence of interest. The result of this extension is a product containing the second adapter sequence, the complement to the sequence of interest, and the complement of the first adapter sequence.



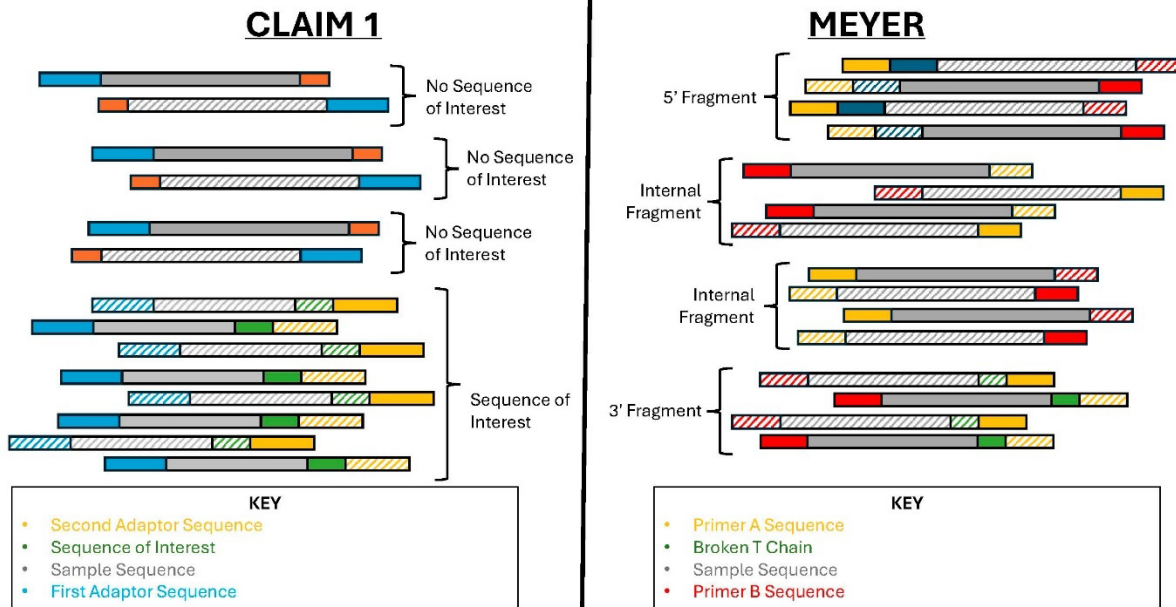
224. No extension products are generated for the fragments lacking a sequence of interest in the claimed methods since there is no site for the target-specific oligonucleotide to target. That is not the same for the Meyer methods. As can be seen, extension products are created for all fragments. For the 5' end, extension products are created that incorporate the Primer A sequence at the end with the cap primer. For the internal fragments, extension products are generated that

contain sequences corresponding to both Primers A and B. For the 3' end, extension products are created that incorporate the Primer A sequence at the end with the cap primer. The only difference between Meyer's general and modified methods is that the extension products of the modified method do not include the cap primer sequence.

225. After extension, amplification occurs. For the claimed methods, only the oligonucleotide extension product sequence corresponding to the sequence of interest and its complement undergo exponential amplification (represented as the dashed lines in the below). For the Meyer methods, all fragments with Primer A and Primer B sequences and their complements undergo exponential amplification (represented as the dashed lines in the below), including fragments that do not contain the sequence of interest.



226. As a result, by the end of the claimed methods, a practitioner would have an enriched library of amplicons that contain the sequence of interest and the necessary adapter sequences for sequencing. These amplicons contain either the oligonucleotide extension product sequence or its complement. The fragments not containing a sequence of interest have not been amplified and do not contain the necessary adapter sequences for sequencing. This is in contrast to the Meyer method where by the end of amplification, there are amplicons for the 5' end, internal fragments, and 3' end fragments (as well as the complements of these fragments).



227. Because Meyer does not disclose the claimed process for selectively amplifying the nucleic acid fragments with a sequence of interest over the nucleic acid fragments without a sequence of interest so to enrich for the sequence of interest, Meyer fails to disclose the claimed amplification process.

2. Any Enrichment Disclosed by Meyer Occurred Prior to Fragmentation

228. IDT and Dr. Sims blur together different parts of the Meyer protocol in an effort to mischaracterize the cDNA fragment amplification process as an enrichment process. *See* EX1085.0098. The amplification of the cDNA fragments does not lead to any enrichment of a target sequence of interest. Dr. Meyer confirms this. EX2287.¶ 75.

229. There are two types of enrichment that occur prior to the first step of the claimed methods. Meyer discloses that they used the SMART PCR cDNA Synthesis protocol to reverse transcribe RNA into cDNA and then to amplify that cDNA. EX1006.0013; EX1034. As explained in the SMART protocol, “[a]ll commonly used cDNA synthesis methods rely on the ability of reverse transcriptase (RT) to transcribe mRNA into single-stranded (ss) cDNA.” EX1034.0004. However, it was understood in the art that this reverse transcription could not “always transcribe the entire mRNA sequence, [so] the 5’ ends of genes tend to be under-represented in cDNA populations.” *Id.*; EX2287.¶ 25. The SMART protocol was designed to address the under-representation of end fragments by “preferentially enrich[ing] for full-length cDNAs.” EX2287.¶ 35. The SMART protocol does this in two ways.

230. The first type of enrichment caused by the SMART protocol is for mRNA transcripts over other types of RNA. Meyer discloses that they isolated and purified total RNA from coral samples. Total RNA refers to the complete set of RNA molecules in the coral samples, including ribosomal RNA. Meyer is interested specifically in the coding regions of the coral (i.e. the transcriptome), so Meyer focuses only on mRNA. Meyer enriches for mRNA over other RNA through the use of the polyA-tail-specific cap primer since other types of RNA found in the sample

lack such tails. EX1006.0012-13; EX2287.¶ 24. Figure 1 of the SMART protocol illustrates this process:

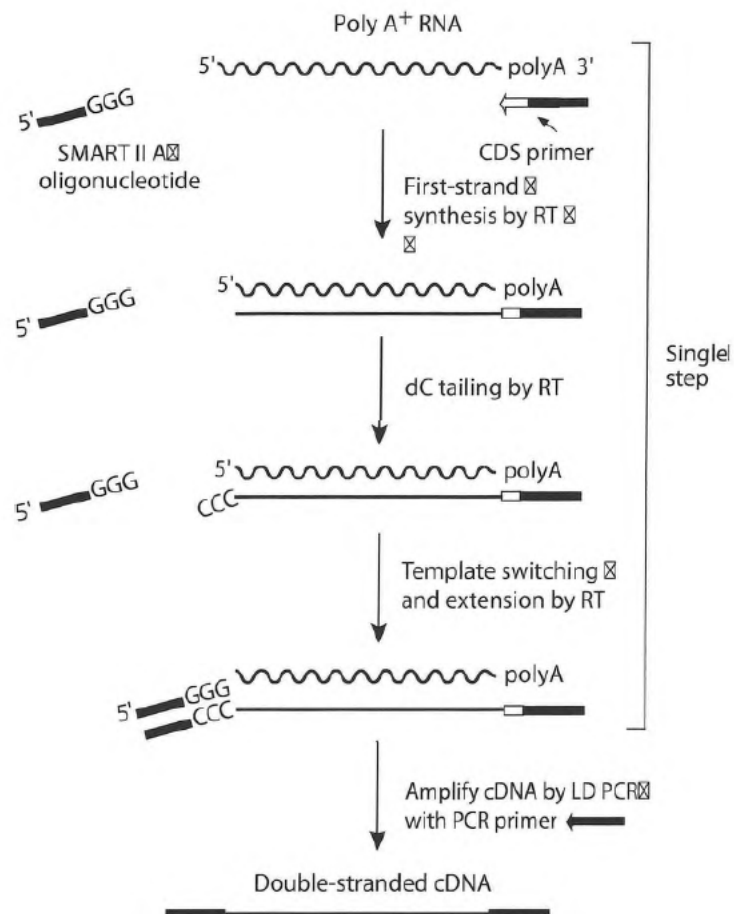


Figure 1. Flow chart of SMART™ technology. The SMART II A Oligonucleotide, 3' SMART CDS Primer II A, and 5' PCR Primer II A all contain a stretch of identical sequence (see Section II for complete sequence information).

EX1034.0005. As can be seen, the cap primer “CDS primer” targets the polyA tail of the mRNA. EX2287.¶ 24. Meyer discloses that they used a customized version of this primer to break the corresponding polyT chain. EX1006.0019; EX2287.¶¶ 31, 67. After annealing this cap primer to the polyA tail, the SMART protocol

discloses that a first strand of cDNA is generated through reverse transcription starting at the CDS primer. By targeting the polyA tail, the SMART protocol enables the enrichment of mRNA over other RNA. EX2287.¶ 24.

231. The second enrichment occurs as part of this same process too. After the first strand of cDNA is synthesized, a second cap primer (“SMART II A Oligonucleotide”) is ligated to the 5’ end of the first cDNA strand and extended. The first and second extensions occur as a single step, which generates full-length cDNA strands that correspond to mRNA. Afterwards, these strands are amplified using primers specific to the cap primers, enriching for the full-length strands. EX1006.0013; EX2287.¶¶ 35-37.

232. By the end of the SMART protocol and before fragmentation, Meyer discloses that they had “preferentially amplif[ied] longer molecules and thus enrich[ed] for full-length transcripts.” EX1006.0013. This conclusion is reinforced by a research paper authored by Dr. Matz that is cited in the next sentence as footnote 41. In that research paper, Dr. Matz explained how the SMART protocol allowed for mRNA populations to be “fully representative,” which he described as “contain[ing] all types of messages originally present in RNA, even the least abundant ones.” EX2074.0003. As I noted above, Meyer couples the SMART protocol with the Evrogen Trimmer Kit to normalize and enrich for full-length transcripts over shorter transcripts. *See supra* 177.

233. At no point does the general method disclose enriching a nucleic acid sequence of interest in a nucleic acid fragment (neither does the modified method).

3. Meyer Does Not Propose Targeted Enrichment of the Broken-T Chain.

234. Meyer explains that the broken chain T-primer sequence is used to “improve reproducibility and further reduce the occurrence of adaptor concatenation.” EX1006.0019; EX2287.¶ 47.

235. Reproducibility is improved because this modified method shortens the fragment length by excluding portions of the homopolymer tail and the cap primer. EX2287.¶ 49. While the 454 sequencer was able to sequence larger fragment lengths than other NGS platforms, the error rate increased as the length of the fragment increased. The modified method excises portions of the 3’ end fragment that are known and not of interest (i.e., the homopolymer tail and cap primer) so that the portions sequenced are reflective of the mRNA sequence. *Id.*, ¶¶ 49-50. By excising the cap primer sequences, Dr. Meyer explains that they would have been able to save 23 nucleotides per fragment, which could have reduced the number of nucleotides that Meyer would have needed to sequence by 13.8 million. *Id.* A POSA would appreciate that considerable cost could be saved by removing unnecessary adaptor sequences from the amplicons to be sequenced.

236. Concatenation, as used by Meyer, refers to primer dimer binding, where the adaptors ligate to each other rather than the transcript fragments. EX2021.¶¶ 58,

86; EX2287.¶ 47. Repeated PCR amplification can exacerbate this issue. The primer dimer complexes contribute to sequencing noise, leading to an increase in read artifacts. By reducing the number of cap primer sites, Meyer reduces the risk of concatenation.

B. The Purpose of Meyer Reinforces that it Does Not Disclose Enrichment

237. The fact that Meyer does not disclose the sequencing of an enriched nucleic acid sequence of interest is not surprising given the purpose of Meyer. Meyer discloses a “method for deep transcriptome sequencing in a single 454 sequencing run” where “[t]he single 454 sequencing run employed in this study produced more than 600,000 sequencing reads, allowing the identification of ~11,000 genes and ~30,000 SNPs.” EX1006.0012. This disclosure means that the method of Meyer allowed for the authors to cast a broad net sequencing as many reads as possible. That is the exact opposite of targeting a sequence of interest and enriching for just that sequence.

238. I understand from Dr. Meyer’s declaration that the amplification methods for the cDNA fragments were not designed to enrich for the 3’ end fragments over cDNA fragments not containing the broken-T chain. EX2287.¶ 65. As Dr. Meyer explains, they were interested in sequencing and assembling the entire coral transcriptome. *Id.* Because they were able to enrich for full-length cDNA strands, there was no need to develop a process to enrich for the 3’ end fragments

over the other cDNA fragments. *Id.*, ¶ 66 And in fact, their process did not. *Id.* As Dr. Meyer explains, their concern with the 3' end fragments being underrepresented in the sequencing data was based on the homopolymer optical issue. *Id.* They fixed that issue by breaking the polyT chain through the substitution of one “T” nucleotide for a “C” nucleotide. *Id.*, ¶ 67 That substitution did not enrich for anything. It only reduced an issue with pyrosequencing. *Id.*

239. I understand that IDT argues that Meyer discloses a method for sequencing an enriched nucleic acid sequence of interest through the targeting of the broken chain T-primer sequence disclosed in Additional File 4. Paper 1, 45. To claim that Meyer discloses any sort of enrichment through that process represents a fundamental misunderstanding of Meyer. Meyer discloses that one major challenge in sequencing the entire transcriptome in one 454 pyrosequencing run is that the 454 sequencing platform suffers from calibration issues caused by homopolymer regions. *See supra* ¶ 88. This issue was well known in the art at the time. *See supra* ¶ 88-89.

240. While any sequence could have homopolymer regions, Meyer focused specifically on the polyA tail at the 3' end of the coral transcriptome. *See supra* ¶¶ 167, 169. Meyer discloses a fix for the 454 sequencing of the polyA tail: disrupting the homopolymer chain “by simply interrupting the poly-T region with a single C.” *See supra* ¶ 169. This interruption created what Meyer described as the “broken

chain T-primer.” EX1006.0009. According to Meyer, “[t]he purpose of the ‘broken chain’ T-primer is to reduce read artifacts during 454 pyrosequencing, which may get thrown out of calibration by too strong of a light signal produced from a long mononucleotide stretch (such as polyT or polyA).” *Id.*, .0019.

241. Further, as Dr. Meyer explains, one reason Meyer uses the SMART protocol is to correct for a potential *over-representation* of 3’ ends caused by reverse transcription errors. EX2287.¶ 25. A POSA would understand the use of this protocol to be strong evidence that the Meyer methods were not designed to enrich for the 3’ end fragment. Instead, consistent with the disclosures in Meyer and Dr. Meyer’s declaration, a POSA would understand that Meyer used the SMART protocol to promote full representation across the transcriptome and not to bias specific regions.

C. The Data in Meyer Does Not Support IDT’s Claims

242. Meyer evaluates the breaking of the T-chain in the general method by comparing the raw reads obtained through the general method with what Meyer expected to see. In the “Results” section, Meyer discloses that a “cDNA sample was prepared from coral larvae as illustrated in Figure 1, and sequenced using the 454 GS-Flx platform. This single sequencing run produced 628,649 reads, with an average sequence length of 232 bases.” EX1006.0002. In the “Discussion” section, Meyer discloses that a “single copy of this ‘broken-T’ adaptor was found in 98,276

of the raw reads (16% of the total), slightly more than the expectation (~10%) based on average read length (232 bp) and assuming an average transcript size of 2,200 bp.” *Id.*, .0009.

243. Although Meyer does not expressly state that raw read data specific to the broken-T adaptor is the same raw read data reported earlier in the publication with respect to the general method, a POSA would have understood that Meyer was referring to the same data set. For example, the average read length of 232 bp reported here is the same as the average sequence length Meyer reported in the “Results” section. Dr. Meyer confirms this understanding. EX2287.¶ 57.

244. Meyer then compares the number of raw reads containing the broken chain T-primer sequence to Meyer’s expectation. EX1006.0009. As disclosed by Meyer, that expectation assumes that the average transcript size is 2,200, and that with an average read length of 232 bp, each transcript would have been fragmented into approximately 10 fragments, where the 3’ end fragment is one of ten. *See id.*

245. As explained above, the modified method uses a primer that is complimentary to the broken chain T-primer, but Meyer discloses the use of the broken chain T-primer in both the general method and the modified method reported in additional file 4. *See supra* ¶¶ 169, 174, 187-188. It appears that IDT conflates these two and mistakes the raw read data that Meyer reported from using the general method as somehow demonstrating enrichment of the 3’ end fragment through the

use of the modified method. Dr. Meyer confirms that this is an error and that the data cannot be interpreted as IDT attempts to do. EX2287.¶¶ 73-75.

246. A POSA reviewing Meyer and considering its data would understand that there are several key limitations. These are all limitations that IDT ignores.

247. First, as explained, Meyer reports the raw read data from using the general method, which does not use an oligonucleotide complementary to the broken chain T-primer sequence for generating extension products. Meyer fails to disclose any raw reads from using the modified method. *See* EX1006.0019 (noting that “the version of the protocol shown here includes minor modifications from that used for preparing coral larval cDNA”); *id.*, .0002 (noting the “cDNA sample was prepared from coral larvae as illustrated in Figure 1,” “sequenced using the 454 GS-Flx platform,” and that “[t]his single sequencing run produced 628,649 reads, with an average sequence length of 232 bases”). Dr. Meyer confirms this. EX2287.¶ 64.

248. Second, the disclosure does not show enrichment. Meyer makes clear throughout the publication that when a homopolymer region is sequenced, the 454 platform can be thrown off calibration, leading to read artifacts. *See supra* ¶ 167. Read artifacts are errors in the sequencing data that can reduce the reliability of the sequencing results. Reducing the number of read artifacts increases the number of reads that can be reliably attributed to specific regions and genes of the transcriptome. That increase, however, does not mean that there are more sequences

in the sequencing population than there might have otherwise been. It only means that more of the sequences already in that population are properly represented in the sequencing data. In other words, the relative population of the sequences remain the same.

249. That is true here. Meyer reports that they fixed the homopolymer optical issue of the 454 sequencer by “simply interrupting the poly-T region with a single C.” EX1006.0009. That nucleotide substitution does not cause enrichment. Instead, it reduces the 454 optical errors associated with homopolymers, allowing for Meyer to identify more of the resulting sequencing reads as 3’ end reads than may have otherwise been possible. Dr. Meyer confirms this. EX2287.¶ 75-76.

250. This concept is reflected in Meyer’s discussion of singleton reads. A substantial fraction of singletons were determined to be “artifacts of cDNA synthesis, normalization, or sequencing.” EX1006.0010. Sequencing artifacts often include sequencing reads that have been erroneously sequenced. Just because the sequencing read cannot be identified does not mean that the amplicon in the sequencing library did not exist. To the contrary, a POSA would have understood that an amplicon could be present in the sequencing library but not reflected in the sequencing data. Rather, that amplicon may be registered as a singleton that likely could not be included in transcriptome assembly. Additionally, because Meyer is performing transcriptome assembly, it is possible that there are sequencing reads that

contain erroneous sequences due to sequencing issues, such as the homopolymer issue, that are still able to be included in the contigs and assembly due to overlapping regions with other singletons.

251. Third, the data cannot be used to infer enrichment. A POSA would understand that the 10% estimate is based on the average sequencing read length and an average transcript length (with no explanation from how that average was obtained). A POSA would understand that the 10% is not based on any prior sequencing results. In fact, a POSA would understand that Meyer does not disclose any prior sequencing runs nor any data from any study to show what percentage of 3' end fragments were observed in sequencing data not using the Meyer methods. Meyer also does not disclose the number of 3' end fragments in the solution before exponential amplification.

252. A POSA would also understand that Meyer fails to control for several independent variables, which all can impact the sequencing results. For example, Meyer made several improvements at once to the sequencing process: (1) Meyer broke the polyT chain to minimize optical issues with sequencing homopolymers; (2) Meyer directed sequencing so that it would start at the internal regions of transcripts and not at the termini; and (3) Meyer fragmented the cDNA strands into fragments approximately 300-400 nucleotides in length, which had been shown to improve sequencing read quality. Meyer also enriched for mRNA transcripts and

full-length cDNA strands to ensure full representation of the transcriptome. Any one of these improvements could have impacted the sequencing read data, and in fact, all of them likely did. A POSA would have understood that it would be improper to assign any specific numerical impact to any one of these fixes given the lack of controls. And as Dr. Meyer recognizes, this is why they used the 10% estimate as nothing more than a directionally gauge to determine whether the breaking of the T-chain worked. EX2287.¶¶ 75-77.

253. Additionally, there is other data missing from the reference that a POSA would have considered critical to making any reliable inferences. For example, Meyer does not disclose the percentage of 5' end fragments in the pre-amplified mixture of cDNA fragments nor in the sequencing data. This data would be critical, especially for the general method, if a POSA were to examine whether the targeting of the cap primer with a complementary oligonucleotide resulted in enrichment. Instead, Meyer only discloses a 10% estimate, which as I explain below, is flawed for several reasons.

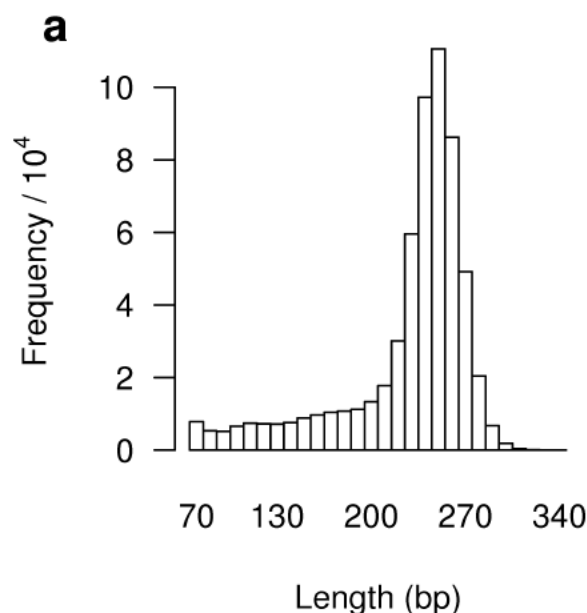
1. A POSA Would Have Understood that the Average Sequencing Read Length Used to Calculate the 10% Estimate Was Skewed.

254. Meyer used the average read length of 232 nucleotides as the numerator for his 10% estimate. As Meyer explains, this average is of the raw sequencing reads. EX1006.0002. Meyer calculated the standard deviation of this average to be

55 reads with a range of read lengths ranging from 30-474. *Id.* Dr. Meyer confirms this. EX2287.¶ 57.

255. Standard deviation refers to the amount of variation in the values of a sample as compared to its average. For a normal distribution (i.e., the standard or “Bell” curve distribution with an equal number of observations above and below “average” or “mean” value), ~2/3 of the data is within 1 standard deviation of the mean, 95% within 2 standard deviations, and 99% within 3 standard deviations. A higher standard deviation value indicates that the values are more spread out from the average than compared to a lower standard deviation value. Here, if the data were normally distributed, the standard deviation might suggest that approximately two-thirds of the raw reads fall between 177 and 287 nucleotides. But a POSA would have understood that the data was skewed.

256. Figure 2(a) of Meyer discloses the “[s]ize distribution of 454 sequencing read sizes after removal of adaptor sequences and outliers.”



EX1006.0004. This distribution reflects the removal of outliers (unlike the raw read data), but even then, it reflects a left-skewed distribution. A POSA would have understood that a left-skewed distribution like this means that the average is less than the median (i.e., the middle value if all values were to be ordered from lowest to highest). A POSA would also understand that a left-skewed distribution means that the median is less than the mode (i.e., the most frequent value in the data set).

257. For left-skewed distributions, a POSA would understand that the median value better reflects the central tendency (i.e. true middle) of the distribution of the data than the average. This is because the median is less affected by the outliers shown in the left tail of the distribution, which result in the average being lower than the true middle. As I noted, this distribution is based on a dataset where Meyer removed other outliers. *Id.* A POSA would have understood that the dataset

of the raw data would have likely been even more skewed to the left before the removal of outliers.

258. As such, a POSA would have understood that the 232 nucleotide average used by Meyer as the numerator for the 10% estimate skewed lower than the central tendency and that the median should have instead been used. This is especially true here where Meyer is attempting to calculate the expected number of sequencing reads containing the broken-T chain, which would be contained in longer reads (~300-400 nucleotides) and not in reads and artifacts that were only 30 nucleotides in length.

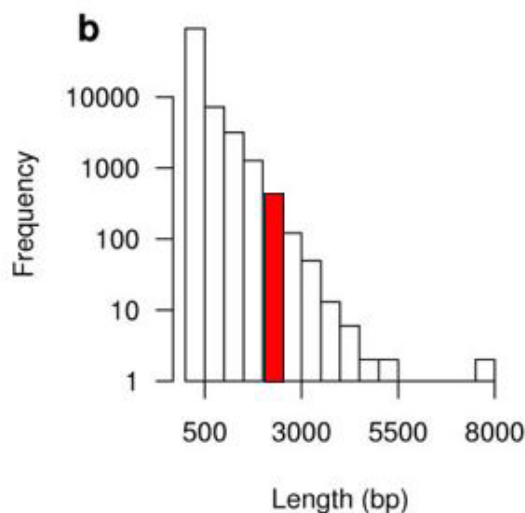
259. While the POSA would not be able to determine the median from EX1006, a POSA would have understood that a more appropriate numerator would have been higher than 232 nucleotides, meaning that the expected percentage would have been greater than 10%. This is one reason why a POSA would have inferred from the calculation that the expected 10% was meant to be directional and could not be used to measure exact changes in proportion. *See* EX2287.¶ 74.

2. A POSA Would Have Understood that the Reported Transcript Lengths Were Likely Skewed

260. As explained above, Meyer relied on an “average transcript length” of 2,200 nucleotides as the denominator for his 10% estimate. Meyer did not disclose from where that average was derived. Dr. Meyer believes that this average was based off the average mammalian transcript length. *Id.*, ¶ 58. If so, a POSA would

have understood that the average transcript length was meant to be directional and could not be used to assess enrichment, especially where the calculation was not reflective of the coral species used in this experiment, or even of invertebrates generally. *Id.*

261. If the average transcript length was instead based on the assembled transcripts after sequencing, a POSA would still have understood the metric to be no more than directional. This is because the distribution data reported by Meyer raises serious concerns as to whether 2,200 bp is a fair average or subjected to outliers that drive the average up. Meyer reported the “[s]ize distribution of assembled sequences after assembly and contig joining” in Figure 2(b):



EX1006.0004 (modified). As can be seen, the distribution data is far from normally distributed. Instead, the data is highly skewed to the right with the modal number of transcripts at the very shortest lengths (e.g., <500 bp bin). In other words, there are

considerably fewer sequences to the right of 2,200 bp (the column reflecting 2,200 bp has been colored red), but the sequences to the right span a greater range of lengths (from 2,200 to ~8,000 bp), thus artificially raising the average over where it would be with a normal or standard distribution. Whereas the average read length skewed left (making the average lower than the central tendency), the average transcript length skewed right, making the average higher than the central tendency.

262. Because this data is not normally distributed, it would be inaccurate to simply rely on the average as the sole measure of “central tendency.” If anything, it is more appropriate to rely on the median (half way point) or mode (most frequent observation). With a right skewed distribution such as this, the median is almost always smaller than the mean. That Meyer did not do so emphasizes how Meyer’s expectation is not meant to reflect a statistically rigorous calculation but instead just a “back-of-the-envelope” guesstimate of what the data might be showing. *See* EX2287.¶ 75.

263. Looking at this distribution, I would expect that on average, the transcript lengths, would be less than 2,200 bp (and maybe even far less). I would also expect the median to be less than 2,200 bp given the skewed distribution (meaning half the transcripts would be less than 2,200 bp). While it is hard to calculate without the full data set, it is not unreasonable to conclude that the median would be less than the mean given the skewed nature of the distribution (consistent

with an exponential distribution). This is key because had Meyer assumed 1,000 bp, then Meyer would have estimated that the 3' end fragments would be present in 23% of reads. Had Meyer assumed 2,000 bp, then Meyer would have estimated that 3' end fragments would be present in almost 12% of reads. Therefore, it is unclear if there is any actual difference between the sequencing reads containing the broken chain T-primer sequence and what *should* have been expected.

264. Transcript lengths can also vary per study, as well as based on the sample source it is isolated from. EX2099.0002. As demonstrated in other publications, the reported average transcript length for *A. millepora* varies (e.g., one recent publication reported that the average transcript length was 1,818 bp). 2018.0004. If Meyer had instead assumed that the average transcript length was 1,818 bp, Meyer would have expected that the 3' end fragment would comprise roughly 13% of the overall read fragments.⁶

⁶ Meyer assumed the average read length was 232 bp based on the raw reads obtained through the general method. To estimate the expected percentage of 3' end fragments in the total raw read data where the average transcript length was 1,818 bp, a POSA would perform the following calculation (consistent with how Meyer did):
$$1/(1818/232) = 1/7.8362 = 0.1276.$$

265. Because of such variation in the average transcript length—and the impact on how many fragments are in one transcript—a POSA would have been cautious to not overextrapolate as to the meaning of any comparisons of that expectation to actual data. This is especially true where neither the data distribution for the numerator (average read length) and the denominator (average transcript length) were normally distributed. A POSA would have understood that this uncertainty is why Meyer describes the 16% of total raw reads containing the broken-T adaptor as just “slightly more than the expectation.” *See* EX1006.0009.

3. If Dr. Sims’ Inferences about the Average Transcript Size are Accurate, then a POSA Would Have Concluded that the Estimated Number of 3’ End Reads Should Have Been Approximately 16%, Not 10%

266. Dr. Sims’ opines that a POSA would have understood that the average transcript length disclosed in Meyer was based off an estimate obtained during the cDNA synthesis step of the methods. Meyer discloses that as part of the cDNA synthesis process, they used gel electrophoresis to determine whether the prepared cDNA transcripts were of appropriate length. *Id.*, .0021. Dr. Sims opines that the 2,200 nucleotide estimate came from this analysis. EX1085.¶ 184 (The POSA would understand that the assumed average transcript size of 2,200 base-pairs was empirically derived from the gel transcript size analysis described at step c of the ‘cDNA amplification’ method at EX1006 (Meyer) at 0021, in which the authors

teach evaluating the length of the cDNA products prior to proceeding with the rest of the method”).

267. This is at best a guess from Dr. Sims, but it is not unreasonable. Dr. Meyer believes that this number was based off of the average mammalian transcript length, which had been known in the art for decades at that point. EX2287.¶ 58. The truth is that a POSA would not have known for certain from where Meyer obtained the average transcript length. However, it is possible that Meyer did obtain that number from the gel electrophoresis run before fragmentation of the cDNA strands. If that is the case, a POSA would have understood that Meyer used the wrong numerator to calculate an estimated percentage of 3’ end fragments.

268. Meyer used the average sequence read length as the numerator. This number is not necessarily the same as the average fragment length. Sequence read length can differ from fragment length for various reasons, including sequencer capabilities, sequencing errors, and other errors that occur as part of the preparation of the sample. EX2003.0005, 0007. EX2087.0001-02. These factors can lead to the sequence read length being less than the fragment length.

269. As such, if Meyer did use the average transcript size as determined before the cDNA strands were fragmented, then a POSA would have understood that the appropriate numerator would have been the cDNA fragment length. Meyer reported that they fragmented the cDNA strands to yield cDNA fragments that were

approximately 300-400 nucleotides in length. EX1006.0013. In Additional File 4, Meyer clarified that the fragments could be between 350-400 nucleotides in length. EX1006.0019. Meyer does not disclose an exact average length for the fragments. But even by the approximate range, a POSA would have understood that the expected percentage of cDNA fragments that contained the broken-T chain should have been roughly 14-18% ($300/2200 = 0.136$; $400/2200 = 0.182$). In fact, using the middle of the range on EX1006.0013 and the bottom on EX1006.0019 (350 nucleotides) results in an estimated 15.9% of cDNA fragments containing the broken-T chain. That is 0.3% **higher** than the actual number of sequencing reads observed by Meyer that contained the broken-T chain (15.6%).

270. Putting this altogether, a POSA would be cautious in what inferences they would have drawn about the number of fragments containing the broken-T chain. A POSA would not have viewed the approximately 6% difference between the actual number of sequencing reads containing the broken-T chain and the expected number to mean that the 3' end fragment was enriched. This is because Meyer did not report adequate baseline numbers for which a POSA could compare the results, and Meyer did not control for the various changes in the general method. A POSA would have understood that the 10% estimate was meant to be directional and not based on the rigorous analysis necessary to assess enrichment. A POSA would have understood that the numerator was likely too low and the denominator

was likely too high based on skewed distributions, and that if the values were normalized, the expectation would be higher. If the average transcript length were based on the length of the cDNA strands before fragmentation, a POSA would have understood that the expected number of 3' end fragments would be between 14-18% based on the approximate cDNA fragment length. As such, a POSA would have understood Meyer's estimated "10%" to be nothing more than a ball-park estimate for Meyer to confirm that their sequence read percentages were in the neighborhood of what they expected. A POSA would have understood that nothing about enrichment could be inferred from these numbers. *See* EX2287.¶ 75.

271. All together, Meyer discloses methods for sequencing the entirety of a transcriptome in one sequencing run and obtaining sequencing results representative of the sequences. Neither the general method nor the modified method of Meyer disclose a process for enriching nucleic acid fragments with target sequences of interest.

272. A POSA's understanding that the Meyer methods do not permit target-specific enrichment of nucleic acid fragments with sequences of interest would be further confirmed by Meyer's SNP analysis. As explained above, Meyer used a new software (at the time) to predict how many SNPs were in the sequencing data. *See supra* ¶ 197. Meyer validated these predictions by targeting 20 SNPs. *Id.* To target these SNPs for sequencing, Meyer used target-specific primer pairs, which were

recorded in Additional File 3. EX2079. Meyer reports on the number of validated SNPs, but Meyer does not compare any target specificity or enrichment from this validation process to the general and modified methods.

273. A POSA would infer from this disclosure that the Meyer methods were not designed to enrich for target-specific sequences. A POSA would understand that Meyer used target-specific primer pairs instead of the Meyer methods because those methods could not be used for targeted enrichment through amplification (especially as claimed by the patents). A POSA would consider that understanding as confirmed by Meyer's failure to apply or even compare the SNP validation enrichment method to the cDNA library preparation methods. This is consistent with Dr. Meyer's understanding of his work. *See* EX2287.¶¶ 73-77.

D. “wherein the nucleic acid fragment comprises a nucleic acid sequence of interest” (’012 patent) / “wherein the nucleic acid fragment comprising the nucleic acid sequence of interest” (’108 patent)

274. Another key deficiency of Meyer is that it does not disclose a “nucleic acid fragment” or a “nucleic acid sequence of interest” as required by the claims of the ’012 and ’108 patents.

275. The “nucleic acid fragment” must contain a “nucleic acid sequence of interest.” The plain language of claim 1 of the ’012 patent defines the “nucleic acid fragment” as “compris[ing] a nucleic acid sequence of interest.” EX1001.0033. The same is true for claim 1 of the ’108 patent. EX1002.0033. A POSA reading these

claims would understand that if there is no sequence of interest, then there is no “nucleic acid fragment” as required by the claims. In reviewing the specification and prosecution history of both patents, I have not seen any disclosures to the contrary. Because the inquiry turns then on whether there is a “sequence of interest,” I have focused my analysis on whether Meyer discloses such a sequence. It does not.

276. I understand that IDT advances a claim construction argument that applies the patents’ definition of “targeted sequence” to “sequence of interest.” Paper 1, 26. For purposes of these proceedings, I have been instructed to analyze “sequence of interest” understanding it to mean “a polynucleotide sequence of interest, for which enrichment is desired.”

277. Under that construction, Meyer fails to disclose a polynucleotide sequence for which enrichment is desired (as there is none). I understand that IDT points to the broken chain T-primer as a “sequence of interest.” *Id.*, 45. A POSA would not have applied that term to the broken chain T-primer because a POSA would have not desired its enrichment.

278. Meyer discloses that “[t]he methods described here for deep sequencing of the transcriptome should be widely applicable to generate catalogs of genes and genetic markers in emerging model organisms.” EX1006.0001. As explained by Meyer, “[t]he characterization of the larval transcriptome for this widely-studied

coral will enable research into the biological processes underlying stress responses in corals and evolutionary adaptation to global climate change.” *Id.* Meyer expresses no interest in the polyA tail of the transcriptome, aside for the problem it causes in sequencing the 3’ end fragments of the transcriptome: “Because 454 sequencing does not efficiently process homopolymer regions greater than 8 bp in length, the polyA tails at the 3’ ends of intact transcripts present a problem, and would be expected to result in under-representation of the 3’ ends of transcripts.” *Id.*, .0009. In other words, Meyer’s focus on the polyA tail is born out of the issues it causes in sequencing the 3’ ends of the transcriptome.

279. As with Meyer, a POSA seeking to perform deep sequencing of the entire transcriptome in one sequencing run would have no interest in enriching the polyA tail, the reverse transcribed polyT tail, or even the broken chain T-primer. The polyA tail is a non-coding region of the transcriptome, meaning that there are no associated genes or genetic markers. *See supra* ¶ 45. Based on the reasons disclosed by Meyer for using either the general or modified methods, it would appear that Meyer is interested in sequencing every other region of the transcriptome but the polyA tail.

280. This view is shared by IDT’s expert witness, Dr. Sims. In explaining how the SMART protocol works, Dr. Sims notes that the “standard Cap Primer contains at its 3’ end a polyT sequence complementary to the polyA tail of the

mRNA, ending in a ‘degenerate nucleotide,’” which is used “to ensure that the Cap Primer only facilitates primer extension when it anneals to the interior 5’ end of the polyA tail, thereby omitting most of the polyA tail.” EX1085.¶ 148. Dr. Sims explains that “[b]ecause the sequence of the polyA tail is not of interest and can also cause sequencing errors, it is desirable to minimize the extent to which it is included in the resulting cDNA.” *Id.*

281. It appears that this recognition that the polyA tail is “not of interest” is the reason why Dr. Sims instead discusses the “sequence regions near 3’ polyA tails.” *See id.*, ¶144. Expanding the “sequence of interest” to apply to the sequence regions of the 3’ end fragment, however, destroys the distinction that the patents make between a “nucleic acid fragment” and a “sequence of interest.” Instead, it would treat the “sequence of interest” as the “nucleic acid fragment.” That is contrary to what the ’012 and ’108 patents disclose. *See supra* ¶ 117, 119. As such, a POSA would not have considered the 3’ end fragment to be a “sequence of interest” as claimed in the patents.

282. It is also important to consider the context in which the patents refer to “sequence of interest.” Both patents are directed to methods for sequencing. A POSA would understand that the desire to enrich for a specific sequence is so that the targeted sequence can then be sequenced and its sequencing data used for some purpose. Meyer’s lack of interest in the sequencing of the polyA tail is apparent by

the cleaned sequencing reads that Meyer deposited in the NCBI's SRA database. *See* EX1006.0016 (disclosing that the cleaned sequencing reads were deposited in the SRA database with accession number SRA003728). Meyer deposited 301,457 reads into the database. Of those reads, only 1,747 reads contain the full broken chain T-primer sequence (CGCAGTCGGTACTTTTTTCTTTTTT)—less than 1% of the total clean reads. A POSA reviewing this contemporaneous data would not have understood Meyer to be interested in the sequencing of the broken chain T-primer sequence when more than 99% of the deposited clean reads are to other sequences.

E. Meyer Fails to Disclose the Other Limitations of the Independent Claims

283. As explained above, each limitation of the independent claims of the '012 patent and '108 patent is conditional on the presence of a “sequence of interest,” a “nucleic acid fragment,” and/or an amplification step that results in enrichment of the sequence of interest. Because Meyer fails to disclose these claim elements, Meyer fails to disclose any of the limitations of the independent claims.

F. It Would Not Have Been Obvious to Modify Meyer to Practice the Missing Claim Limitations

284. I understand that IDT argues that Meyer renders obvious the independent claims. IDT's arguments, however, do not address modifications of Meyer that would address the specific missing limitations identified herein. Because Meyer does not disclose the “amplify ... to enrich” or “sequence of interest”

limitations, as explained above, and IDT does not point to any other prior art that would cure these defects, I disagree with IDT and Dr. Sims that the challenged independent claims are obvious.

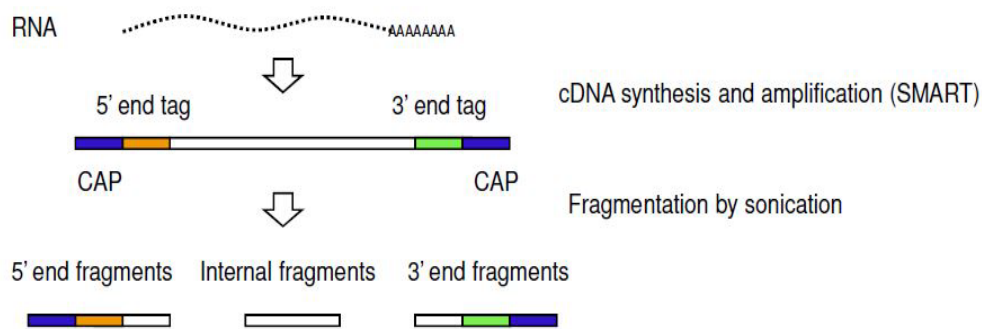
285. Because the independent claims are not obvious over Meyer, it is my understanding that the dependent claims cannot be either.

G. Meyer Fails to Anticipate or Render Obvious Claim 21 of the '012 patent

286. Claim 21 of the '012 patent is directed to the method of claim 1 and further requires that “the nucleic acid fragment of the nucleic acid fragment ligated to the partial duplex adaptor in step a) is obtained by performing reverse transcription on an RNA fragment prior to obtaining the nucleic acid fragment ligated to the partial duplex adaptor in step a).”

287. A POSA would have understood that this claim limitation requires the fragmentation of RNA prior to reverse transcription. *See supra* ¶¶ 46, 157-160. A POSA would have also understood that “an RNA fragment” as used by the claim language refers to the product of the fragmentation process. *Id.*

288. Meyer discloses that the RNA is reverse transcribed using the SMART protocol before it is fragmented.



EX1006.0013. Meyer does not disclose an RNA fragment as it would have been understood by a POSA. Rather, Meyer discloses that mRNA (not RNA fragments) are reverse transcribed into cDNA and that cDNA is fragmented by sonication, resulting in cDNA fragments.

289. IDT and Dr. Sims claim that the mRNA used in Meyer are “fragments.” As explained above, a POSA would not have understood the mRNA that is reverse transcribed in Meyer to be products of a fragmentation process. *See supra* ¶¶ 169-170. It would also have not been obvious to a POSA how to practice the method of claim 1 by fragmenting the mRNA before reverse transcription based on Meyer’s disclosures given how reliant Meyer is on using the SMART protocol in their methods. I understand that IDT and Dr. Sims provide no explanation as to why a POSA would have been motivated to make such a modification or how a POSA would do so. It is my opinion that a POSA would not have been motivated to do so with a reasonable expectation of success.

290. A POSA would have understood that this ordering was important to the Meyer methods. Meyer sought to enrich for full-length cDNA strands through the SMART protocol, which is a protocol for reverse transcribing mRNA and then enriching for longer, full-length cDNA strands. A POSA would have understood that fragmenting the mRNA transcripts before using the SMART protocol would have prevented the subsequent enrichment for full-length cDNA strands.

291. A POSA would have also understood that it would not be desirable to fragment the RNA before reverse transcription given that it risked the unwanted biasing of amplification products during the cDNA synthesis process, which would negatively impact Meyer's ability to sequence fragments corresponding to the entire transcriptome. This is reflected by Meyer's use of the Evrogen Trimmer Kit and SMART protocol to counteract the amplification bias towards shorter transcripts. EX1006.0013; EX2287.¶ 23.

292. Rather, a POSA would have understood that without serious revisions (none of which are disclosed in Meyer), fragmenting the RNA before reverse transcription would have likely resulted in many sections of the transcriptome missing from the sequencing data. A POSA modifying the Meyer method would not have been motivated to fragment the RNA before reverse transcribing given that it would require the POSA to develop an entirely new cDNA synthesis method that capped only the termini of the mRNA strands, enriched for full-length cDNA, and

incorporated the broken-T chain into the 3' end fragment cDNA. Rather than ensuring that the transcriptome was fully represented in the data, a POSA would have understood that fragmentation before reverse transcription in the Meyer method would increase the level of under-representation in the data for various segments of the transcriptome. Fragmenting the RNA before reverse transcription would have also prevented a POSA from using the cap primers disclosed in Meyer. Without these cap primers, Meyer fails to disclose a method of controlling for the primer sequence that would be added to the termini of the original RNA transcripts. There is no reason that a POSA modifying Meyer would alter it to fragment the mRNA strands first.

293. Additionally, a POSA would not have had a reasonable expectation of success in modifying Meyer to fragment the mRNA before reverse transcription given that Meyer fails to teach how to enrich for mRNA fragments over other RNA fragments and for full-length cDNA strands without the use of the cap primers.

H. Meyer Fails to Anticipate or Render Obvious Claim 22

294. Claim 22 of the '012 patent is directed to the method of claim 1 and further requires that “each end of the nucleic acid fragment is ligated to a partial duplex adaptor.”

295. It is my understanding that IDT and Dr. Sims concede that Meyer does not describe the ligation of partial duplex adapters to each end of the end fragments,

which IDT claims to contain a sequence of interest. In fact, Meyer discloses that the cap primers are intended to block ligation of the partial duplex adapters at the termini of the cDNA strands. EX1006.0019; EX2287.¶¶ 42-45.

296. Instead, it is my understanding that IDT and Dr. Sims rely on an accident that may happen as a result of fragmentation but that is not a guaranteed outcome. Dr. Sims claims that during fragmentation, some cap primers are fractured. Partial duplex adapters can be ligated to these damaged cap primers.

297. While it is possible that some cap primers are damaged by fragmentation, a POSA would have understood that it was not a guaranteed outcome. Whether any cap primer is damaged during fragmentation depends on a number of variables, including how long fragmentation occurs, the desired size of the fragments, the number of cDNA strands present in the solution during fragmentation, and even the method used for fragmentation.

298. Meyer discloses that in their experiments, they sonicated cDNA for a series of increasing durations from 1 minute, 3 minutes, 5 minutes, and 10 minutes. EX1006.0023. Meyer teaches that the treatment that generates fragments from about 300 to about 1,000 nucleotides should be used. *Id.*

299. Assume that Dr. Sims is correct that the average transcript length is based on the full-length cDNA strand before fragmentation. A POSA would have understood that using a lower duration treatment that yields fragments that are about

1,000 nucleotides in length would result in roughly one break per strand. Meyer discloses that the cap primer sequence is 23 nucleotides in length. *See id.*, .0019. That means that per strand, there are approximately 2,177 nucleotides that are not part of the cap primer, meaning that approximately 99% of the potential locations for the single fracture are not part of the cap primer.

300. The Meyer method does not require that any specific number of cDNA strands undergo fragmentation. Meyer discloses that the “methods described here for deep sequencing of the transcriptome should be widely applicable to generate catalogs of genes and genetic markers in emerging model organisms.” *See id.*, .0001. A POSA would have understood that the method as described in Figure 1 (or even Additional File 4) could be used with a smaller number of mRNA strands. The smaller the population of cDNA strands that are fragmented, the more likely the situation that none of the cap primers are damaged.

301. A POSA would also understand that the selected fragmentation method could impact the outcome. For example, Meyer teaches that nebulization can be used instead of sonication. Sonication and nebulization fragment by different mechanisms. Sonication randomly disrupts DNA molecules in solution while nebulization fragments DNA molecules by forcing them through a small hole in the nebulizer unit. EX1009.0023, EX2088.0001. As such, nebulization does not randomly fragment DNA strands the same way as sonication. *Id.*

302. Because of the options available under the Meyer methods, it cannot be said that fragmentation with the Meyer methods would inevitably result in damaging the cap primers. It also cannot be said that any damage sustained by the cap primers would inevitably lead to the ligation of a partial duplex adapter to the damaged cap primer. While possible, there are instances where neither occurs.

303. I understand that Dr. Sims identifies a small subset of reads that may contain a broken cap primer and that may have a partial duplex adapter ligated to their end. *See* EX1085.0166-0168. It is notable that the amount of 3' end sequencing reads that Dr. Sims identified is less than 2% of the 98,276 raw reads reported by Meyer. This underscores that while there is a possibility that damage may occur, it is not inevitable in every instance of the method.

304. Additionally, while Dr. Sims relies heavily on the sequencing reads deposited in the NCBI Sequencing Database (*see* EX1085.0124), there are certain unknown variables that raise questions as to what inferences can be drawn from the sequences. First, Dr. Sims reports that the database contains 301,457 individual reads. EX1085.¶ 181. Meyer notes that “[t]he cleaned sequencing reads produced in this study have been deposited in NCBI’s SRA database” but fails to explain the criteria for the cleaned sequencing reads. EX1006.0016. For example, Meyer reports 599,248 sequencing reads after adapter trimming and size-selection. That is almost double what is contained in the SRA database. Nowhere does Meyer report

what specific steps were used to generate 301,457 sequencing reads. A POSA would have been cautious to make inferences based on the deposited data without understanding how the data was cleaned.

305. I understand that IDT and Dr. Sims make no arguments with respect to how it would have been obvious to adapt the Meyer method to practice this claim. In my opinion, it would not have been obvious.

306. A POSA would not have been motivated to modify Meyer to allow for ligation of the partial duplex adapter to each side of the end fragments given that Meyer expressly blocks such ligation as a means of controlling for which primer sequence is located at the termini. There would have been no benefit to the Meyer methods to modify them so that a partial duplex adapter could be ligated to both ends.

307. A POSA would not have had a reasonable expectation of success in modifying Meyer to allow for ligation of the partial duplex adapter to each side of the end fragments. Meyer relies on the cap primers to control for sequencing read directionality. Meyer fails to disclose a method for otherwise ensuring that all termini have the correct primer sequence since Meyer's PCR suppression method only blocks amplification of fragments with identical flanking primers. Meyer does not teach another way to ensure the Primer B sequence is located at the interior and not termini.

IX. Claim 2 is not obvious over Meyer in View of Siebert

308. I understand that IDT argues that Meyer could be adapted to practice dependent claim 2 of the '012 patent. Paper 1, 72. I also understand that IDT makes this same argument for claim 2 of the '108 patent. My opinion is the same with respect to both patents. Claim 2 requires that the “nucleic acid fragment” of claim 1 comprises genomic DNA. As explained above, a POSA would not have been able to practice claim 1 of the patents based on the disclosures and teachings of Meyer. Siebert does nothing to remedy that issue.

309. Siebert is missing numerous key disclosures. For example, Siebert fails to disclose any method for sequencing the genomic DNA, especially with NGS technology, which would not come into existence for nearly a decade after Siebert. *See* EX1007.0002. The Siebert method is a library preparation method used for gel electrophoresis. Siebert uses gel electrophoresis to estimate the size of various genes, which they then cross-reference to known genetic sequences in an attempt to map out specific regions.

310. As I explained above, this process is commonly referred to as “genome walking.” *See supra* ¶¶ 62-66. It was known to be a laborious process that focused on small, very specific regions of the genome. A POSA would have understood that while the methods of Meyer and Siebert both concern genetic material, the methods are in many ways opposites of one another. Where Meyer seeks to sequence the

entire transcriptome, Siebert is mapping a few genes in the human genome. Where Meyer is interested in the coding regions of the coral larva (as evidenced by Meyer's enrichment of mRNA transcripts over total RNA), Siebert is interested in studying both coding and non-coding regions of the genome.

311. Both references fail to disclose or teach the claimed method of selective amplification. As explained above, Meyer discloses the amplification of fragments corresponding to the entire length of the cDNA strands, failing to teach an amplification method that enriches for cDNA fragments containing the sequence of interest. *See supra* ¶¶ 214-227. Siebert instead discloses a method where fragments containing both universal adapter sequences are suppressed and not amplified through the use of a “special adaptor.” EX1007.0001. Unlike the claimed methods where amplification is driven by both universal adapter sequences, Siebert discloses that amplification is driven by “defined, distal, gene-specific primers.” *Id.* Siebert discloses that the “high specificity of the method” is “primarily due to the suppression effect of the adaptor/adaptor primer combination.” *Id.*, .0002. A POSA would have understood that Siebert is attributing the specificity to the use of vectorette adapters and PCR suppression, where any fragment containing the universal adapter sequences will be suppressed and only extension products generated through the gene-specific primer will be amplified.

312. I understand that IDT claims that a POSA would have been motivated to combine the two references because of “the similar chemistries of cDNA and genomic DNA.” Paper 1, 69. I find that argument overly simplistic. While cDNA and genomic DNA share the same nucleotide bases, that argument ignores the proper inquiry here, which is not whether a POSA would have been motivated to prepare and sequence genomic DNA but instead whether a POSA would have been motivated to modify the Meyer methods to do so. A POSA would not have been motivated to do so for several reasons.

313. First, how Meyer obtains and uses the cDNA would have discouraged a POSA from adapting the method to be used with genomic DNA. This is because Meyer relies on a specific protocol (the SMART protocol) to target the polyA tail of mRNA, convert the mRNA to cDNA, and use the cap primer sequences installed during cDNA synthesis to enrich for full-length cDNA strands. In the case of genomic DNA, a POSA would be left without the reverse transcription process, which is key to the Meyer methods. Without considerable alteration, a POSA would only be able to practice the Meyer methods by translating genomic DNA into RNA and then reverse transcribing that RNA back into DNA (i.e., cDNA). A POSA would not have been motivated to do so when methods existed for ligating adapters to the ends of genomic DNA fragments.

314. Second, a POSA would also not have been motivated to do so where, in the case of Siebert, the POSA is interested in sequencing both coding and non-coding regions of the genome. cDNA only contains sequences for the coding regions. A POSA would not be able to sequence non-coding regions through the use of cDNA. To practice the Meyer methods would require the POSA to ignore sequencing information for the non-coding regions of the genome, which a POSA would not have been motivated to do.

315. Third, a POSA would not have been motivated to adapt Meyer for use with genomic DNA given that the purpose of Meyer is to address a specific sequencing issue with the 454 sequencer for mRNA sequences. As explained above, Meyer introduces the broken-T chain into cDNA strands as a means of correcting for an optical sequencing error caused by the sequencing of homopolymer regions. *See supra* ¶ 169. This fix is specific to the polyA tail of mRNA. There is no direct comparison for genomic DNA. A POSA would have little reason to take a method fixing an issue with RNA polyA tails and adapt it for use with genomic DNA.

316. Fourth, a POSA would not have been motivated to use Siebert to modify the Meyer methods to practice the claimed inventions. The methods described by Siebert predate NGS by nearly a decade. As Dr. Sims testified, a POSA would not have looked to these older methods to develop a method specific to use with NGS. EX2037.0003-04 (152:14-153:2). But even had a POSA looked to Siebert, it would

have found Siebert to be discouraging. Siebert teaches away from amplification with single target-specific primers where no other safeguards are available. For example, Siebert reports that “both ‘unpredictably primed PCR’ and ‘vectorette PCR’ generated multiple PCR products even after nested PCR was performed.” EX1007.0001. As Siebert explains, “the presence of multiple PCR products would complicate further characterization without Southern blot hybridization or extensive cloning.” *Id.* A POSA would have understood Siebert to be reporting on the presence of off-target products, which would not have been desirable in a target enrichment scheme. Siebert teaches that the way to minimize these off-target products is through the use of “special adaptor[s]” that combine vectorette PCR with PCR suppression techniques. Siebert confirms that these additional safeguards are what led to target specificity, not the use a single gene-specific primer.

317. In fact, Siebert published another paper the next year in which Siebert explained that even with these methods, off-target specificity remained an issue that was only exacerbated through PCR amplification. EX2043.0001. As I explained above, these concerns are repeated over and over again in the various genome walking publications. *See supra* ¶ 66. Well before 2012, a POSA would have understood that these publications taught away from the use of a single target-specific primer without the addition of many complicated steps. *Id.* With this

understanding, a POSA would not have been motivated to modify Meyer to practice the claimed inventions.

318. I understand that IDT claims that a POSA would have had a reasonable expectation of success given “the past successful use of anchored PCR with genomic DNA and/or Siebert.” Paper 1, 72. I disagree. A POSA would not have expected that Meyer could be adapted to practice the claimed methods. This is because of the decades of publications in the art that demonstrated how the use of a single target-specific primer without additional safeguards would not work. *See supra* ¶ 66.

319. Additionally, Meyer fails to teach how to target genomic DNA. Meyer discloses that through the SMART protocol, mRNA can be enriched from total RNA. Meyer does not offer any disclosures as to how to do the same for genomic DNA. Additionally, Meyer relies on the use of cap primers to target and identify end fragments. As explained above in my discussion of why Meyer does not disclose nor teach a method for fragmenting RNA before reverse transcription, Meyer fails to teach any methods for capping one end of specific fragments where there are other fragments in the solution. *See supra* ¶¶ 294-307. At best, Meyer discloses how to target mRNA and to cap full-length strands. It would not have been clear to a POSA how to do this for genomic DNA, which would have been excised from the genome prior to further manipulation.

320. Siebert teaches that PCR suppression and special adapters drive target specificity. A POSA would not have had a reasonable expectation of success in adapting Meyer where the POSA would be using universal adapter sequences and would not be relying on PCR suppression.

321. It is my opinion that IDT and Dr. Sims have failed to explain why a POSA would have reasonably expected to adapt Meyer to practice the claimed methods.

X. Claims 17-19 of the '012 patent and Claims 11-12, and 19 of the '108 patent are not obvious over Meyer in view of Bronner and/or Caruccio

322. I understand that IDT argues that Meyer could be adapted to practice dependent claims 17-19 of the '012 patent. Paper 1, 75. I also understand that IDT claims that Meyer could be adapted to practice dependent claims 11-12 and 19 of the '108 patent. For the reasons explained below, I disagree.

323. For the '012 patent, claim 17 requires performing bridge amplification before sequencing the enriched amplicons generated by the method of claim 1. Claim 18 requires using four labeled reversible terminators as part of the sequencing that occurs in the method of claim 1. I agree with Dr. Sims that both methods are specific to Illumina NGS platforms. *See* EX1085.0192. Therefore, my analysis considers whether it would have been obvious for a POSA to adapt the Meyer methods for use with Illumina sequencing platforms. Claim 11 of the '108 patent is also directed to the use of four labeled reversible terminators. Claim 19 is directed

to a flow cell, which I also understand to be Illumina specific. I understand IDT's arguments to be the same across these four claims. I address them together.

324. Conversely, claim 19 of the '012 patent requires using semiconductor sequencing with the method of claim 1, and therefore requires the use of Ion Torrent NGS platforms for the claimed methods. I agree with Dr. Sims that this form of sequencing is specific to Ion Torrent NGS platforms. *See* EX1085.0197. Claim 12 of the '108 patent also concerns the same limitation. As such, my analysis considers whether it would have been obvious for a POSA to adapt the Meyer methods for use with Ion Torrent.

325. It is my understanding that IDT and Dr. Sims argue it would have been obvious to a POSA in light of Bronner and/or Caruccio that Meyer could be adapted for use with the Illumina and Ion Torrent NGS platforms. I disagree for several reasons.

326. First, neither Meyer, Bronner, nor Caruccio disclose library preparations similar to the claimed inventions. Bronner describes a general PCR amplification method in its "Basic Protocol 4," which ligates adapters to all fragments in a sample, after which the fragments are amplified via PCR from the adapter. EX1009.0015; *see also id.*, .0012. Indeed, the only disclosure of gene-specific primer is in an "alternative protocol," and even this method uses two target specific primers. *Id.*, .0017-18. Specifically, Bronner discloses that a pair of gene-

specific primers with tailed universal primers could be used to amplify specific fragments. *Id.* Neither protocol teaches the use of a single target-specific primer to selectively amplify target sequences and thus enrich for those sequences.

327. Similarly, Caruccio teaches a method where transposase enzymes (enzymes which can move DNA sequences) are used to fragment and introduce adapters to DNA in a randomized process. EX1008.0001. This technology is the same Nextera technology that is disclosed in Grunenwald, which is expressly described as facilitating random fragmenting and adapter integration. EX2056.0001 Caruccio fails to teach the use of target-specific primers, let alone a single target-specific primer, or the use of selective amplification to enrich for target sequences.

328. Second, a POSA would not have been motivated based on Bronner and/or Caruccio to adapt the Meyer methods to be used with Illumina or Ion Torrent. as explained above, Illumina sequencers made up more than half of the market by 2012 and were perceived as superior to the 454 sequencer. *See supra* ¶ 82.

Indeed, Bronner itself recognized that “the Illumina SBS sequencing by synthesis approach was the dominant sequencing technology.” EX1009.0001. Further, a POSA would have also understood that the 454 and Illumina platforms were distinct methods using different chemical methods to conduct sequencing. *See supra* ¶ 89. There is no reason that a POSA would have been motivated to take a paper focused on 454 library preparation methods paper specifically focused on improvements to

fix 454 sequencing issues, such as Meyer, and adapt the methods for Illumina. *See id.* This is especially true where a POSA would have understood that Illumina systems did not suffer from the homopolymer issues Meyer sought to address, due to the nature of Illumina's sequencing chemistry. *See id.*

Additionally, in the event a POSA wanted to modify a method for Illumina sequencer, that POSA would have considered Illumina library preparation methods as the more effective starting point than a 454 library preparation method. *See supra* ¶¶ 81-82.

329. Similarly, there is no reason a POSA would have looked to a dated 454 sequencer method to develop an Ion Torrent library preparation method. I understand that IDT argues that a POSA would have been motivated because Ion Torrent and the 454 sequencer both use emulsion PCR and well-based semiconductor sequencing platform. Paper 1, 79. However, Ion Torrent and the 454 sequencer employ distinct sequencing detection methods. *See supra* ¶¶ 84-88, 91. The 454 sequencer employed an optical detection method, while Ion Torrent utilized a pH sensor. *Id.* Meyer describes a method specifically designed to fix issues with the optical detection method caused by homopolymers. While Ion Torrent was also recognized to have limited sequencing capacities with respect to homopolymers, there is no reason that a POSA would have assumed a correction for an optical detection method would have been applicable to the Ion Torrent's pH sensors. *See*

EX2086.0001-02, EX2098.0008. Additionally, POSA would have understood that there were differences in the library preparation methods with Ion Torrent and 454 such that it would have been easier to start from an Ion Torrent sequencing preparation method rather than a 454 library preparation method. *See supra* ¶¶ 84-88, 91-92.

330. I understand that IDT argues that a POSA “would have been motivated to switch from a 454 sequencer to a semiconductor sequencer to save costs.” Paper 1, 79. While cost certainly would have been a factor to a POSA in considering which sequencing platform to use, the choice of sequencer is not the same as choosing whether to adapt a library preparation method from an old sequencer for a new sequencer. Rather, a POSA would start by considering library preparation methods specifically designed for the sequence they wanted to use, and then consider how to optimized those methods. Indeed, Caruccio is consistent with this approach, discussing discrete library preparation methods for different sequencers rather than couching the methods as universal to the different sequencing platforms. *Compare* EX1008.0003, 0006 (Roche/454-Compatible Library Production and Preparation) to *id.*, .0004, 0009 (Illumina/Solexa-Compatible Library Production and Preparation). Simply put, by 2012 a POSA seeking to design an optimized library preparation method would have more likely started with publications directed to Illumina or Ion

Torrent sequencing library preparation methods rather than a dated 454 sequencing library preparation method.

331. Additionally, Bronner and Caruccio both teach away from elements of the library preparation method claimed in the '012 and 108 patents. Bronner teaches methods for avoiding PCR bias, stating that while “PCR amplification during library prep allows lower amounts of starting material to be used, [it] introduces bias.” EX1009.0019. Bronner describes polymerases that “amplify complex fragment libraries with the least amount of bias” and stating that “[e]xcessive PCR cycling should be avoided as it can lead to bias in the library.” *Id.*, .0015. Bronner further discourages the use of target-specific primer pairs during library preparation, stating that “[u]nnecessary PCR amplification steps will exacerbate amplification biases, and so they should be avoided wherever possible.” *Id.*, 0017. Indeed, Bronner expressly teaches a “PCR-free library method” to circumvent these issues. *Id.* The PCR-free method does not employ any target-specific primers or an amplification scheme to enrich for target sequences. A POSA considering Bronner’s teachings would not have been motivated to modify Meyer to promote target biasing and enrichment through its PCR amplification scheme.

332. Similarly, Caruccio discourages ligating partial duplex adapters to nucleic acid fragments as part of the library preparation process. As described above, Caruccio states that “[u]sing in vitro transposition to simultaneously fragment

and tag DNA in a single-tube reaction is a significant improvement upon current procedures, which generally [have] consist[ed] of distinct DNA fragmentation, end-polishing, and adaptor-ligation steps.” EX1008.0005. Caruccio criticizes those procedures as “often involv[ing] significant sample loss and hands-on time” while “suffer[ing] from limited sample throughput.” *Id.*, 0001-02. Caruccio proposes that a commercial kit developed by Nextera is a preferable solution to avoid these issues, and that the Nextera kit can “can be exploited to perform all of these steps in a single, 5-min reaction.” *Id.* The Nextera kit expressly bypassed the standard fragmentation, end processing, and adapter ligation workflow to instead randomly fragment DNA and insert transposon ends onto the ends of the fragments. *Id.*, .0006. The transposon ends could then be used to facilitate introduction of universal adaptor sequences at the ends of the fragments. *Id.* As the transposomes introduce transposon ends at both ends of double-stranded DNA fragments, a POSA would have viewed using target-specific primers to incorporate universal primers as redundant to the transposon method, as the transposons “serve as the foundation for subsequent adaptor addition.” *Id.*, 0002. As such, Caruccio’s teachings would not have motivated a POSA to modify Meyer to incorporate the universal primer sequences through the use of a partial duplex adapter and a target-specific primer, as in the method described by the ’012 and ’108 patents. A POSA would also not have

been motivated to use transposomes, which insert at random points in DNA, to target specific sequences of interest.

333. IDT argues that a POSA would have had a reasonable expectation of success in modifying Meyer for use with Illumina and Ion Torrent sequencing platforms in light of the methods described in Caruccio. But there is no reference in any of these references to any method that enriches a target sequence using a single target-specific primer. Further, neither Bronner nor Caruccio provides teachings that would assist a POSA in modifying Meyer's amplification scheme to selectively target and enrich for specific fragments over others.

XI. SECONDARY CONSIDERATIONS

334. The nonobviousness of the claimed methods is reinforced by objective indicia of nonobviousness, including at least that (1) the prior art taught away from the claimed methods, (2) the art expressed skepticism towards the claimed inventions, (3) there existed long-felt, unmet needs that were resolved by the claimed inventions, and (4) the industry praised how the claimed inventions resolved those needs.

A. Nexus

335. As explained above, the claimed inventions are directed to the use of a single target-specific primer (the "one or more oligonucleotides" with "at least 10 bases designed to be complementary to the nucleic acid sequence of interest") to

selectively amplify and enrich a target sequence with a pair of common adapters that can then be sequenced on a massively parallel sequencing platform. *See supra* ¶¶ 116-120. This combination of limitations is unique and characteristic for all claims of both patents. Tecan (and its predecessor, NuGEN) refer to the claimed technology as “Single-Primer Enrichment Technology” or “SPET.”

336. The specification of the patents discloses “[p]roducts based on the methods of the invention may be commercialized by the Applicants under the trade name Ovation®. Ovation® is a trademark of NuGEN Technologies, Inc.” EX1001.0031 (31:30-33); EX1002.0031 (31:30-33). Example 1 of the patents demonstrates the claimed methods through the use of NuGEN’s Ovation Ultralow Library System. EX1001.0031; EX1002.0031.

337. After the priority date, NuGEN published an advertisement in Nature’s Blood Cancer Journal where NuGEN described the technology as such:

SPET (single-primer enrichment technology) another innovative NuGEN technology, maximizes the information yield from NGS with a rapid and efficient sample-preparation workflow. SPET is a target-enrichment tool that allows for the simultaneous detection of single-nucleotide polymorphisms (SNPs), copy-numbers variants (CNVs) and insertions or deletions. Until recently, CNV and sequence analysis would have required completely different analytical platforms and sample preparation to elucidate information for these three types of genetic variation. SPET does so in a single assay, conserving patient samples and making more efficient use of sequencing resources. SPET is the core technology

used in Ovation Cancer Panel 2.0 Target Enrichment System.

SPET can also be used to identify gene-fusion events and sequence variants in an RNA sample. The SPET technology for targeted RNA sequencing is flexible enough to target any gene in any organism for gene-fusion detection and gene-expression analysis. This will be an important contribution to RNA-seq-based cancer diagnostic and prognostic tests.

EX2082.0001

338. In 2015, named inventors Jonathan Scolnick and Douglas Amorese reported on the use of SPET. EX2083.0001. In the publication, they illustrate how the Ovation Target Enrichment System utilizes SPET.

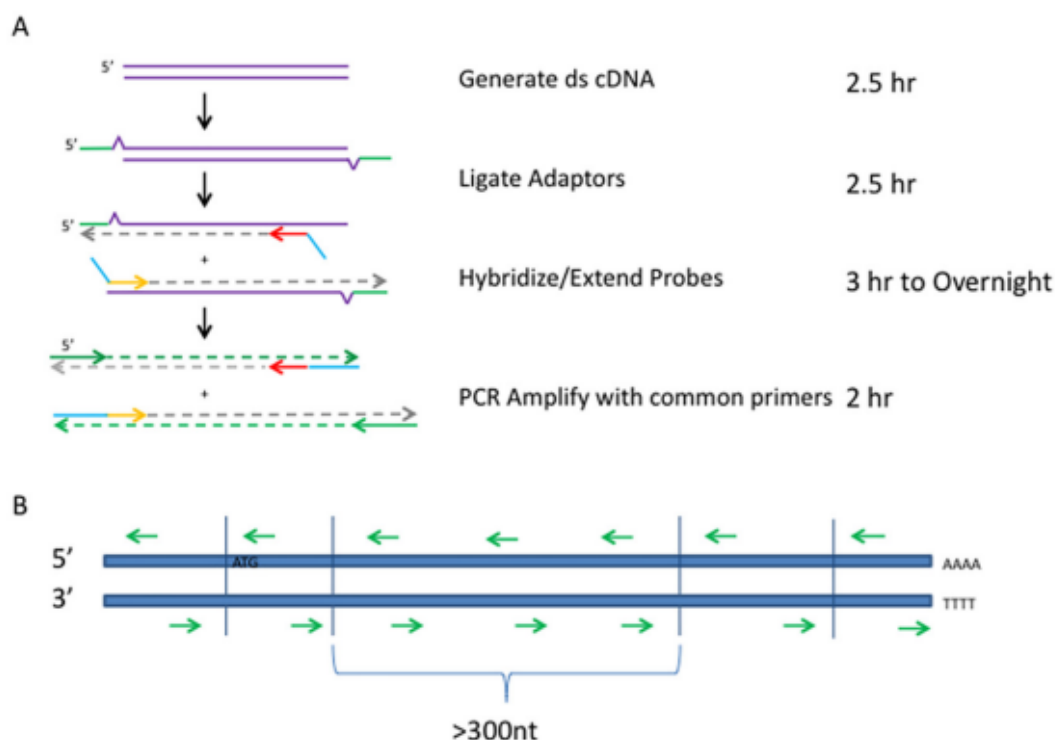


Fig 1. Description of the Ovation Target Enrichment System. (A) Experimental steps of the assay and time required for each step. Adaptors (green) are ligated on to generated double stranded cDNA (ds-cDNA). Probes (shown in red and yellow) are hybridized to target cDNA and extended with a polymerase (dashed grey lines). All probes have common tail sequences (blue), which are used as priming sites along with adaptor sequences in subsequent library amplification PCR steps. (B) Example of probe positioning across different exons in a full length double stranded cDNA. Each exon (demarcated by blue vertical lines) will have probes (green arrows; arrow points in the 3' direction) designed to hybridize near the predicted exon-exon junctions. Exons larger than 300 nucleotides (nt) may have additional probes tiled along the length of the exon to obtain more complete sequence coverage. Probes are designed against both strands of the cDNA to enable identification of gene fusions when only one of the pair of genes is targeted. Translation start sequence (ATG) and poly A tail are labeled.

doi:10.1371/journal.pone.0128916.g001

Id., .0002.

339. This illustration is coextensive with steps 1(a) through 1(f) of the '012 patent and 1(a) through 1(c) of the '108 patent. The publication explains that the “resulting library was sequenced on an Illumina MiSeq,” satisfying the last steps of both independent claims.

340. In 2017, NuGEN announced the issuance of U.S. Patent No. 9,650,628 (“the ’628 patent”). The ’628, ’012, and ’108 patents all claim priority to the same provisional application. *See supra* ¶ 135. NuGEN explained at the time that the ’628 patent describes “the Single Primer Target Enrichment Technology (SPET) which enables targeted nucleic acid analysis by next generation sequencing (NGS). NuGEN offers commercial solutions using SPET to create target and strand-specific libraries ready for NGS.” EX2084.0001.. NuGEN stated at the time that “[p]roducts that SPET is currently incorporated [include] Ovation® Custom Target Enrichment System, Ovation® Custom Gene Fusion Target Enrichment System, and Allegro Targeted Genotyping by Sequencing (TGBS).” *Id.*

341. After Tecan acquired NuGEN, Tecan incorporated SPET into the Allegro® Targeted Genotyping V2. According to its product sheet, it “us[es] the patented Single Primer Enrichment Technology (SPET) approach to specifically target SNPs of interest.” EX2085.0001. It notes that this technology provides for “[h]ighly scalable multiplexing to enable largescale genotyping.” *Id.*

342. The Allegro Targeted Genotyping V2 product sheet also depicts its workflow:

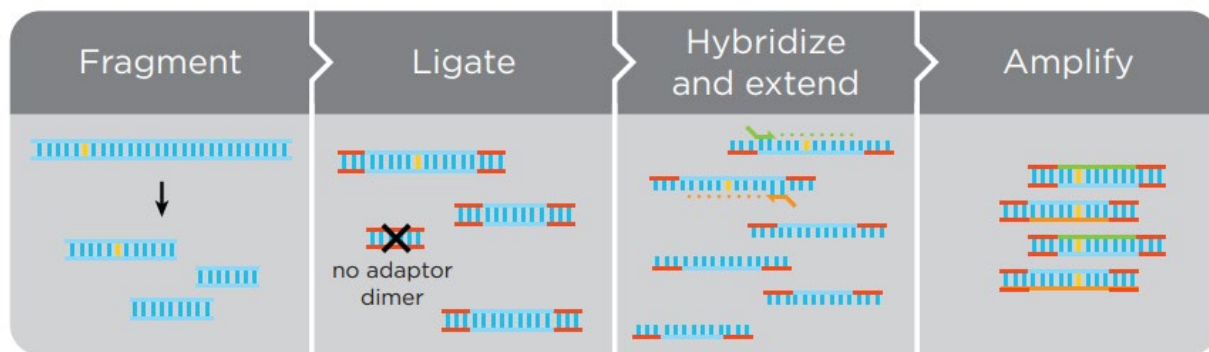


Figure 1: Allegro Targeted Genotyping V2 is a complete end-to-end solution for targeted genotyping, with a simple easy to follow workflow.

Id.

The product workflow demonstrates that the product is coextensive with the claimed inventions. Nucleic acids are fragmented, and the nucleic acid fragments are ligated to partial duplex adapters. An oligonucleotide complementary to a nucleic acid sequence of interest is annealed to the sequence of interest and extended. These extension products are amplified using primers directed to the adaptor sequences to generate libraries of amplified products containing the oligonucleotide extension product sequence or its complement. As explained on the product sheet, this library preparation method generates amplicons that can then be sequenced using NGS.

Id., .0001-02.

343. Based on the patents, the prior statements made by Tecan and NuGEN, and the product manuals illustrating the commercial embodiment of SPET, it is my conclusion that the Tecan Ovation and Allegro products embody the claims of the '012 and '108 patents.

B. Teach-Away/Skepticism

344. The methods claimed in the '012 and '108 patents disclose the use of a single target-specific primer to generate oligonucleotide extension products containing a sequence of interest and the amplification of the oligonucleotide extension products to generate enriched sequences of interest that can be sequenced through NGS. In my opinion, as of 2012, a POSA would have been aware of the nearly two decades worth of literature that demonstrated concerns with amplification, such as target specificity and off-target binding. *See supra* ¶¶ 62-67. Additionally, a POSA would have been aware from literature concerning the development and improvement of NGS library preparation methods that the need for accurate target specificity was paramount. *See supra* ¶¶ 93-111.

345. As described in the background section, the prior art taught away from the use of a single target-specific primer, especially for amplification, given concerns with specificity and off-target binding. *See supra* ¶¶ 62-67. For example, some of the earliest target-specific primer genome walking methods were developed expressly to address issues with “[i]llegitimate products result[ing] from non-specific annealing of either primer,” which was characteristic of the earliest vectorette PCR reactions. EX2044.0001.

346. Indeed, Siebert recognized these issues as early as 1996—a year after the Siebert prior art reference relied on by IDT published. Siebert noted that many single target-specific primer genome walking methods suffered from “a high

background of non-specific PCR products,” which often made “it necessary to do a second round of PCR with nested primers to achieve greater specificity.” EX2043.0002. Siebert explained that using PCR suppression could improve on the technique; however, Siebert also acknowledged that the amplification of some cDNA still yielded high levels of unwanted amplification products. *Id.* In those instances, Siebert recommended the use of additional, more complex methods to filter out background products. *Id.*, .0006.

347. Similar observations appear in the art only a few years later in 1998, when Padegimas *et al.* attempted to address the “low specificity, low efficiency, and/or extensive cycling that contributed to a higher error rate in PCR products” in then-current PCR based walking methods, including vectorette, splinkerette, and adapter-based methods. EX2045.0001. Notably, the single target-specific primer method advanced by Padegimas required both a non-target degradation system and extensive optimization to achieve a reduction in off-target product generation. *Id.*, .0003-05.

348. Despite these efforts, more than seven years later Miura *et al.* disclosed that “single unique primer in combination with a common primer” genome walking methods were still understood to “often [be] plagued with non-specific amplification products derived from so-called ‘off-targets.’” EX2047.0001. Further, Miura understood that this off-target phenomena was characteristic of single target-specific

primer techniques rather than conventional PCR amplification methods, which “use[] a pair of unique primers to selectively amplify the target sequence” that achieve specificity “by a synergistic effect between the specificities of the two primers.” *Id.*

349. Nearly five years later, Bowers *et al.* noted that “[d]espite claims of specific amplification of the desired [] region of a target gene, a common problem that occurs in RACE reactions is obtaining non-specific amplification products.” EX2049.0002. Similarly, a year later, in a patent filing, Sohn *et al.* reiterated that while “many studies have been made to improve the genome walking method,” “it is still in need of improvement.” EX2050.0003. Sohn stated that the “most serious problems of the PCR-based genome walking methods are low amplification efficiency and non-specific amplification.” *Id.* Sohn explained that to “improve this problem, complicated multi-step methods were developed, including secondary PCR, purification of desired PCR products using biotinylated primers, or improvement of PCR cycle conditions.” *Id.* Even then, “satisfactory performance [had] not yet been achieved.” *Id.*

350. More broadly, the art taught away from and expressed skepticism at various amplification methods. Tsoktouridis *et al.* conducted a comparison of various PCR amplification methods and found “no simplicity, low reproducibility, high backgrounds of nonspecific PCR products, insufficient homopolymer tailing,

small PCR product amplification, and laborious time-consuming procedures.”

EX2048.0002. Indeed, prior to development of NGS methods, PCR was known to introduce issues into genetic sequence analysis due to errors and amplification bias.

EX2025.0001, 0006. For example, Syvänen (a 2001 review article concerned with different methodologies for SNP genotyping that pre-date the advent of NGS platforms) discusses the use of allele-specific primers in PCR (a method of genetic analysis using PCR). EX2025.0001. Syvänen discloses that using such primers “without a consecutive target-specific detection reaction or size-separation step, the specificity of the method might be hampered owing to primer dimers and other spurious amplification products that will not be distinguished from actual PCR products.” *Id.*, .0006. Syvänen notes that a “limitation of all variants of allele-specific PCR is that the reaction conditions or primer design for selective allele amplification must be optimized empirically for each SNP,” and as such, “are best suited for the analysis of a limited number of SNPs in large sample collections.” *Id.* Syvänen discloses that the technology “should avoid the PCR amplification step, and the technology could be based on the analysis of single DNA molecules, which would allow direct determination of the haplotypes formed by the SNPs.” *Id.*, .0011.

In a table where Syvänen identifies the “most significant advantage,” “disadvantage,” or “special feature” of various SNP-genotyping methods, Syvänen recognizes that the most significant advantage of one technology is that “PCR

amplification [was] avoided” and that a special feature of another is that it “avoids target amplification.” *Id.*, .0005.

351. Even with the advent of NGS methods, prior art publications continued to suggest that “the best way to avoid bias during PCR is to avoid library amplification by PCR altogether.” EX2042.0011. As described above, several different groups developed library preparation methods that bypassed PCR amplification. *See supra* ¶¶ 68-69. Indeed, Bronner and Quail expressly taught that “[u]nnecessary PCR amplification steps will exacerbate amplification biases, and so they should be avoided wherever possible” and that it was possible to perform NGS without PCR amplification. EX1009.0018-19. In 2011, Quail reported that “[l]ibraries prepared without PCR amplification performed best.” EX2071.0001. Additionally, as of the early 2010s, various commercial groups had invested time and resources developing PCR-free sample preparation kits for use in NGS. *See supra* ¶¶ 68-69.

352. As such, a POSA would have understood that the field was skeptical of the specificity of single target-specific primer methods due to the generation of off-target products. *See supra* ¶¶ 62-67. A POSA would be particularly skeptical of coupling a single target-specific primer method with an NGS platform, as NGS platforms already suffered from issues related to PCR bias. *See supra* ¶¶ 68-69. Indeed, a POSA would have understood from the art at the time that most single

target-specific primer methods employed complicated experimental parameters designed to reduce off-target primer binding, including complicated vector schemes, chemical modification strategies, and non-target degradation treatments. *See supra* ¶¶ 62-67. Moreover, a POSA would have appreciated that most of those solutions served only to reduce off-target effects rather than eliminate them. *Id.*

C. Long-felt, unmet need

353. Despite the benefits conferred by the massively parallel nature of NGS, researchers quickly recognized that early NGS methods suffered from “the inability to easily target specific regions of a genome.” EX2062.0003. As such, researchers recognized that new “sequence capture methodologies [were] required for time- and cost-effective sequence analysis in high-throughput genomic projects.” EX2064.0001. Such improvements would need to be accurate, efficient, and modular. Indeed, researchers in the field observed that while the “ease of any given enrichment method and its potential for near-term implementation in a production setting are also considerations,” “the quality of the NGS data produced with a given enrichment method on a given sample type will ultimately dictate the strategies we elect to further optimize.” EX2063.0002. Further, early PCR-based target enrichment methods used in conjunction with NGS methods were known to suffer from accuracy issues. *See supra* ¶¶ 68-69, 97-106. For example, Kelley (a reference relied on by IDT) reported that even with primer pairs, they experienced

unintentional amplification of unwanted sequences, which led to them discarding half of their sequencing reads. EX1010.0005 (explaining that the primer pairs used did not have the desired specificity to ensure that only the target sequences were being amplified).

354. Other early NGS-compatible target capture methods relied on specialized instrumentation, such as RNA library “bates” (short RNA molecules designed to selectively pull nucleic acid fragments of interest out of a sample) and microfluidic microarray chips (tiny silicon or glass hybridization arrays), to use, which made the methods technical and complex. EX20692.0004; EX2064.0001-02. Indeed, in some instances, the lack of methods not requiring specialized instrumentation resulted in researchers re-using instruments, including hybridization arrays, despite the risk of sample contamination. EX2051.0004. Researchers recognized at the time that “there is greater demand for cost-effective, streamlined library preparation methods.” EX2056.0001. One such commercial solution was developed by Nextera, which sought to improve library preparation efficiency by streamlining sample preparation steps and limiting PCR amplification. *Id.* However, the Nextera method utilized a nonspecific transposase method that indiscriminately fragmented and tagged nucleic acid sample fragments. *Id.*, .0001-02 As such, the Nextera method was not viable when experimental parameters required a target capture approach.

355. In addition to the lack of accurate, efficient, and user friendly methods, PCR-based target enrichment methods suffered from logistical issues that complicated applying such methods at scale. EX2010.0002. Such methods were viewed as “laborious, time-consuming and costly,” particularly when employed in a multiplexed fashion, where it was “difficult to check successful amplification per fragment.” EX2051.0003. Moreover, such methods increased the “chance of obtaining artifacts.” *Id.* Alternative capture methods, such as capture-by-circularization, suffered from issues with consistent target capture and therefore produced variable sequencing results. *See* EX1001.0006 (2:18-24). Such methods, similar to traditional multiplexed PCR, where each target required two target specific primers or probes, were ill-suited for use with massively parallel NGS methods which could sequence far more nucleotide sequences than could practically be targeted. EX2010.0002

356. As of late 2011, targeted enrichment, including PCR-based methods, capture-by-circularization, and hybrid capture, all suffered from issues with “uniformity, efficiency of coverage and off-target capture.” EX1050.0007; *see also* EX2050.0002-04. Indeed, some researchers concluded that “it may ultimately be more cost-effective to sequence a whole genome” rather than use target capture. EX1050.0007. Researchers noted that target enrichment and sequencing library preparation methods were “laborious, time-consuming, prone to experimental errors

and difficult to automate” and had complex protocols which require large amount of starting genomic DNA.” EX2052.0002. Such methods were further “susceptible to coverage biases secondary to amplification and molecular artifacts.” *Id.*

357. These long-felt, unmet needs are the same reported by the patents: “[t]here is a need for improved methods for selective target enrichment that allow for low-cost, high throughput capture of genomic regions of interest without specialized instrumentation.” EX1001.0016; EX1002.0016 “Additionally, there is also a need for high efficiency nucleic acid library generation.” *Id.* As the patents explain, the “methods of the invention described herein fulfills these needs.” *Id.*

D. Industry Praise for Tecan’s Solution to these Long-Felt, Unmet Needs

358. Over the years, there have been numerous articles published in the industry praising the claimed inventions and how the claimed inventions addressed the above long-felt, unmet needs.

359. Barchi *et al.* is a research paper published in 2019 by researchers at four institutions in Italy and one in Spain. In it, the authors test NuGEN’s SPET technology. Specifically, the authors identify this technology by stating that “[r]ecently Nugen® developed the single primer enrichment technology (SPET, United States Patent 9,650,628).” EX2065.0002. They report using NuGEN’s Allegro® SPET products. *Id.* In the publication, they describe SPET as “a new, robust, customizable solution for targeted genotyping” and “a customizable solution

for targeted sequencing at an affordable price.” *Id.*, .0001-02. They explain that the technology’s “reliance on single primers, rather than on primer pairs, greatly simplifies panel design, and allows higher levels of multiplexing than PCR-based genotyping.” *Id.*, .0001. They conclude by stating that the technology “performed well with DNA samples prepared by different laboratories using different DNA mini-preparation protocols,” that it provided “the ability of multiplexing thousands of samples in a single sequencing run,” and that it “allow[ed] the discovery of thousands of novel SNPs not originally included in the panel.” *Id.*, .0012-13.

360. Caitlin Smith, who obtained her Ph.D. in neuroscience Yale University and frequently authors publications concerning genomics and neuroscience research in Biocompare’s Future Lab,⁷ published an editorial article in 2021 discussing the claimed technology. Smith reports on “target enrichment using single primers, supported by Tecan Genomics’ Allegro® Target Genotyping V2 with Single Primer Enrichment Technology (SPET).” EX2067.0003. Smith describes the technology in this way:

Enzymatically fragmented sample DNA is ligated to adaptors containing barcodes with their DimerFree® technology. The barcoded samples are pooled and then hybridized to probes/primers, which are extended and the results amplified with PCR to create a targeted library.

⁷ Biocompare’s Future Lab is exclusively sponsored by IDT.

Id. Smith explains that “[u]sing single primers instead of primer pairs greatly simplifies the design of primer panels, and allows higher levels of multiplexing compared to amplicon-based enrichment.” *Id.*

361. Scaglione *et al.* is a research paper published in 2019. It reports on a study using SPET technology that was conducted by researchers at four Italian institutions and one British institution. In it, the researchers “evaluate single-primer enrichment technology (SPET) reproducibility.” EX2068.0003. They report on using the Allegro targeted genotyping kit and Ovation Target Enrichment kit, which they explain are “based on the same SPET technology.” *Id.* They explain that “[u]nlike other methods, [the claimed] technique allows for the execution of cost-effective genotyping experiments while providing full control on target sites,” which it attributed in part to the “high efficiency enrichment system” and “the scalability to tens of thousands of probes in a single reaction.” *Id.*, .0006. They report that the study “results suggest that [the claimed] SPET genotyping is a very effective approach for genotyping a large number of loci,” and “suggest[ed] that it might be a valuable tool for large-scale genotyping studies in virtually all species.” *Id.*, .0008.

362. Benjamino *et al.* is a research paper published in 2021. The authors all belong to the Jackson Laboratory for Genomic Medicine, which is a prominent and well-respected research laboratory. In the paper, they reported using the Allegro Targeted Genotyping kit. EX2069.0002. While they report adapting the “single

primer extension reaction that is widely used for genotyping [] and implement[ing] it as a quantitative, straightforward, and cost-effective method for profiling mouse microbial communities,” their method still employs the same patented SPET technology. *See id.*, .0002, Fig. 1. They ultimately described using the SPET technology in this way “as a quantitative, straightforward, and cost-effective method for profiling mouse microbial communities.” *Id.*, .0002.

363. A POSA would have understood from the industry praise that the claimed inventions succeeded in providing an effective method for using single target-specific primers with universal adapters to selectively amplify and enrich sequences of interest that could then be sequenced through NGS, despite the concerns in the art with using a single target-specific primer and PCR amplification either together or even separately.

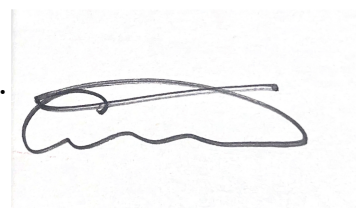
XII. CONCLUSION

364. In summation, Meyer fails to disclose the claimed methods of selectively amplifying for the oligonucleotide extension product sequence or its complement to enrich for a sequence of interest, thus failing to anticipate or render obvious the independent claims of the '012 and '108 patents. Because Meyer fails to anticipate or render obvious the independent claims of the '012 and '108 patents, Meyer necessarily fails to do the same for the dependent claims of the patents. IDT's additional prior art references—Siebert, Bronner, and Caruccio—do not remedy these deficiencies.

365. Tecan's inventions are novel and non-obvious. The named inventors devised the claimed methods despite the weight of the prior teaching away from and expressing skepticism towards the path they took. By inventing the claimed methods, the named inventors addressed several long-felt, unmet needs, which is reflected in the industry praise of the claimed inventions.

366. I declare under penalty of perjury that the foregoing is true and correct.

Executed on July 18, 2025, at Redwood City, California.

A handwritten signature in dark ink, appearing to read 'Carlos D. Bustamante', is written over a light blue rectangular background.

Carlos D. Bustamante, Ph.D.