THE INVENTION OF NEXT-GENERATION SEQUENCING

Caressa N. Tsai†

TABLE OF CONTENTS

DOI: https://doi.org/10.15779/Z38V40K11S

† Ph.D., McMaster University, Department of Biochemistry and Biomedical Sciences, 2021; J.D. Candidate, University of California, Berkeley, School of Law, Class of 2024. I am grateful to Peter Menell and Allison Schmitt for their guidance in preparing this Article. Special thanks to Christine O'Brien Laramy, Hunter Kolon, Will Kasper, Duane Yoo, and the participants of the 2022–23 Life Sciences & Innovation Workshop, for feedback, editorial assistance, and many helpful discussions.

I. INTRODUCTION

In the 1950s, the discovery of the structure of DNA[1] and the theory behind its function as the fundamental blueprint of life[2] launched modern

1. James D. Watson & Francis H. C. Crick, *Molecular Structure of Nucleic Acids*, 171 NATURE 737 (1953) (describing, as one leading example from the 1950s, the double helix structure of DNA); *see also* Rosalind E. Franklin & Raymond G. Gosling, *Molecular Configuration in Sodium Thymonucleate*, 171 NATURE 740 (1953).

2. Francis H. C. Crick, *On Protein Synthesis*, 12 SYMPS. SOC'Y FOR EXPERIMENTAL BIOLOGY 138 (1958) (suggesting that genetic information flows through biological systems

molecular biology. Scientists recognized that DNA is: (1) the genetic code of all living organisms; and (2) composed of two twisted strands of base-paired nucleotides.[3] As soon as the molecular biology community recognized the importance of DNA and its composition, the race to develop DNA sequencing technologies—methods to determine the order of the nucleotides in a strand of DNA—commenced.[4] Today, DNA sequencing is among the most important techniques driving life sciences research, with DNA aptly perceived as the key to unlocking new diagnostic and therapeutic strategies.[5]

Scientists began developing "first-generation" methods of DNA sequencing in the 1970s. This early research led to the invention of Sanger sequencing, which enabled the Human Genome Project (HGP). By the completion of the Project in 2003, the pursuit of "next-generation" DNA sequencing—comprising methods that were faster and cheaper than their first-generation counterparts—had begun in earnest. And in 2006, a Cambridge-based company called Solexa launched one of the first next-generation sequencing (NGS) machines. The next year, Solexa was acquired by Illumina, a company that would go on to dominate the NGS market.

The advent of NGS launched an "omics" era of modern medicine. Omics broadly encompasses all approaches aimed at comprehensively interrogating the "building blocks"[6] of life, primarily: DNA (genomics), RNA (transcriptomics), and protein (proteomics). The omics revolution, fueled by NGS, shifted scientific inquiry from reductionist to holistic strategies.[7] With

---

only in certain directions, two of which being from DNA to RNA, and from RNA to protein, but other directions being possible as well); *see also* Matthew Cobb, *60 Years Ago, Francis Crick Changed the Logic of Biology*, 15 PLOS BIOLOGY e2003243 (2017).

3. *A Brief Guide to Genomics*, NAT'L HUM. GENOME RSCH. INST., https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics (last visited Nov. 24, 2022).

4. *See Frederick Sanger Interview,* NOBEL PRIZE (Dec. 9, 2001), https://www.nobelprize.org/prizes/chemistry/1958/sanger/interview/.

5. Marcos Morey et al., *A Glimpse Into Past, Present, and Future DNA Sequencing*, 110 MOLECULAR GENETICS & METABOLISM 3, 3–4 (2013) (noting that "genetic diagnostics, biotechnology, microbiological studies, forensic biology, and systematic[] . . . taxonomy" have all benefited from NGS development).

6. Relevant to this concept is the central dogma, which explains the flow of genetic information from DNA, to RNA, to protein (or RNA to protein). *See generally* James A. Shapiro, *Revisiting the Central Dogma in the 21st Century*, 1178 ANNALS N.Y. ACAD. SCIS. 6 (2009).

7. *See* Rebecca K. Delker & Richard S. Mann, *From Reductionism to Holism: Toward a More Complete View of Development Through Genome Engineering*, *in* PRECISION MEDICINE, CRISPR, AND GENOME ENGINEERING: MOVING FROM ASSOCIATION TO BIOLOGY AND THERAPETUICS 45, 46–47 (Stephen H. Tsang ed., 2017); *Beyond Conventional Cell and Molecular Biology Research Methods*, ILLUMINA, https://www.illumina.com/areas-of-interest/cellular-molecular-biology-research.html (last visited Sept. 24, 2022) (explaining that traditional methods in molecular biology "seek[] to understand the function of a single gene, gene family, or signal transduction

NGS, scientists could produce "big" data quickly and inexpensively, which ushered in new perspectives in biology. Researchers recognized that biological systems were not discrete units, but complex, networked landscapes, and that phenotypes often resulted from not just individual genes, but full genomic profiles.[8] NGS allowed the scientific and medical communities to approach disease treatment at a previously unfathomable resolution.

This Article explores the development of NGS for DNA studies (genomics) with a focus on the Illumina sequencing platform as the leading technology in this space, and the motivational factors critical for Illumina's success. The discovery story that led to Illumina's ongoing dominance in the NGS market spans multiple countries, companies, universities, and scientists. Over several years, a mixture of factors contributed to the remarkable innovation that resulted in the Illumina NGS platform, clustering into two separate stages. First, scientific curiosity, altruism, public funding sources, academic recognition, and serendipity motivated foundational research. Then, as the Solexa team expanded their idea into a dominating NGS platform, the landscape of "innovation drivers" shifted to private funding sources, patent protection, well-timed licensing, dedication to commercialization potential, and aggressive litigation.

Part II summarizes the foundational "first-generation" technology that inspired NGS development, the technical details of the modern-day Illumina NGS platform, and the modern life sciences applications of NGS. Part III traces the history and development of the NGS platform in five phases, from the use of Sanger sequencing in the HGP to the early 2000s competition between Solexa and other startup companies in bringing the first NGS machine to market. Part IV analyzes the interplay of several innovation drivers that contributed to the Illumina NGS discovery story in two distinct stages. Finally, Part V discusses the state of modern DNA sequencing technologies.

II. TECHNICAL BACKGROUND

The Illumina NGS platform is a complex system, with technical features that can be traced back to innovations in molecular biology research from the

family" while NGS technologies "broaden cell and molecular biology research . . . [and] enable[] analysis across the genome, transcriptome, and epigenome").

8. *See* Delker & Mann, *supra* note 7, at 49; Jeffrey Gagan & Eliezer M. Van Allen, *Next-Generation Sequencing to Guide Cancer Therapy*, 7 GENOME MED. 1, 1 (2015) (explaining the shift in cancer research from reductionist thinking (that all types of cancers developed from individual genetic mutations, minimizing the number of relevant biological actors) to systems-wide thinking (that only some cancers are caused by single mutations, but most are genetically complex and involve dysregulation of multiple *pathways* rather than genes)).

1970s. This Part first describes the foundational processes of *in vivo* DNA replication, polymerase chain reaction (PCR), and Sanger sequencing, and then turns to the technical details of the Illumina NGS platform and the modern-day applications of NGS.

A. FOUNDATIONAL TECHNOLOGY

NGS technologies rely on two foundational inventions: PCR and Sanger sequencing. PCR, Sanger sequencing, and NGS technologies—while remarkably innovative—are also *in vitro*, synthetic mimics of *in vivo* DNA replication, which occurs naturally in all organisms. This Section explains *in vivo* DNA replication, PCR, and Sanger sequencing and highlights shared components between all three processes and NGS (**Table 1**).

Table 1: Key technical components of foundational and modern DNA sequencing technologies.

	***In vivo* DNA Replication**	**Polymerase Chain Reaction**	**Sanger Sequencing**	**Next-Generation Sequencing**
Template	Single-stranded DNA			
Nucleotides	Unmodified nucleotides		Unmodified nucleotides and labeled dideoxynucleotides	Reversible terminator nucleotides
Primers	Primase enzyme	Forward and reverse primers	Sequencing primer	
Enzyme	Polymerase			
Kinetics	Polymerase enzyme adds unmodified nucleotides to produce a complementary strand of the template DNA	Polymerase enzyme adds unmodified nucleotides to produce millions of copies of the template DNA	Polymerase enzyme adds unmodified nucleotides to produce a complementary strand of the template DNA; sometimes will add a labeled dideoxynucleotide instead and induce chain termination	Polymerase enzyme adds reversible terminator nucleotides to produce a complementary strand of the template DNA

1. In vivo *DNA Replication*

In living organisms, DNA exists as a double-stranded helix, with two complementary strands comprised of four types of nucleotides (adenosine (A), thymidine (T), cytidine (C), guanosine (G)) paired together.[9] These strands are also referred to as chains, and are intertwined due to base-pair complementarity between the purine-based (A, G) and pyrimidine-based (T, C) nucleotides.[10] For replication, a protein separates DNA into two single strands, so that each one can serve as a template for a complementary strand.[11] Cellular machinery builds a complementary strand based on base pairing between nucleotides.[12] The entire set of DNA within an organism is called its genome and is organized into chromosomes.[13]

There are four main components required for *in vivo* DNA replication, all of which have analogs in PCR, Sanger sequencing, and NGS: (1) a single-stranded template DNA strand; (2) nucleotides (free A, T, C, G to be added); (3) a primase enzyme (to establish a double-stranded foundation from replication to proceed from); and (4) a polymerase enzyme (to catalyze the addition of each nucleotide to the growing complementary strand).[14] Briefly, the polymerase enzyme attaches to the primase-defined region and physically moves along the template DNA strand, sequentially adding nucleotides to a new strand based on complementarity to bases in the template (**Figure 1A**).[15] The end product is a freshly synthesized, complementary chain of DNA.[16] This replication process starts at many randomly distributed points throughout genomes[17] and ends at similarly distributed points.[18] Therefore, the length of

9. BRUCE ALBERTS ET AL., MOLECULAR BIOLOGY OF THE CELL (2002).

10. *Id.*

11. *Id.*

12. *Id.*

13. *A Brief Guide to Genomics*, *supra* note 3.

14. This is a simplified explanation that omits some of the molecular players in this process. A more comprehensive summary of DNA replication is described elsewhere. *Id.*

15. *Id.*

16. *Id.* Polymerase enzymes specifically catalyze the formation of phosphodiester bonds between nucleotides, linking them together in growing DNA chains. The bond forms between the 3' end (hydroxyl) of one nucleotide and the 5' end (phosphate) of the next nucleotide. *Sanger Sequencing Steps & Method*, SIGMA ALDRICH, https://www.sigmaaldrich.com/US/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing (last visited Sept. 27, 2022) [hereinafter Sigma on Sanger].

17. Michalis Fragkos et al., *DNA Replication Origin Activation in Space and Time*, 16 NATURE REVS. MOLECULAR CELL BIOLOGY 360 (2015).

18. James M. Dewar & Johannes C. Walter, *Mechanisms of DNA Replication Termination*, 18 NATURE REVS. MOLECULAR CELL BIOLOGY 507 (2017).

template DNA to be complemented in each round of replication is indeterminate, but is certainly smaller than the length of the entire genome.[19]

2. Polymerase Chain Reaction

PCR is an *in vitro*, experimental analog of *in vivo* DNA replication.[20] The reagents required for PCR are similar to those for *in vivo* DNA replication: (1) a single-stranded template DNA strand (to be amplified); (2) nucleotides (free A, T, C, G to be added); (3) forward and reverse primers (to establish double-stranded foundations from which amplification can begin); and (4) a polymerase enzyme (to catalyze the addition of each nucleotide to the growing strand copy). The steps of PCR are also essentially the same as those of *in vivo* DNA replication: the polymerase enzyme attaches to the primer-defined regions and physically moves along the template DNA strand, sequentially adding nucleotides to a new strand based on complementarity to bases in the template (**Figure 1B**).[21] However, this PCR process repeats iteratively across several cycles, amplifying the template strand millions of times.[22] External temperature triggers mirror the physiological conditions that help *in vivo* DNA replication proceed.[23] Most importantly, high temperature cycles induce repeated denaturing of double-stranded DNA (the complementary DNA strands that are synthesized are, at first, bound to the original template strand) into the single-stranded form required for the cycle to repeat.

A key difference between PCR and *in vivo* DNA replication is the *target region* of DNA to be amplified (i.e., the boundaries of the template strand). As described in Section II.A.1 *supra*, *in vivo* DNA replication occurs at multiple points throughout an organism's genome.[24] For PCR, researchers may extract the entire composite of genomic DNA from a target organism to use as a template, but focus on a specific target region to be amplified based on the selection of primer sequences (replacing the primase enzyme of *in vivo* DNA replication). These forward and reverse primers face inwards towards each other and define the boundaries of the target template strand to be synthesized.

---

19. *See id.*

20. For a graphic illustration of the PCR process, see *infra* Figure 1B.

21. Elizabeth Pelt-Verkuil et al., *A Brief Comparison Between In Vivo DNA Replication and In Vitro PCR Amplification*, *in* PRINCIPLES AND TECH. ASPECTS OF PCR AMPLIFICATION 9, 12 (2008).

22. *Id.*

23. *In vivo* DNA replication requires the concerted activity of many different proteins and physiological conditions, to maintain growing DNA chains in appropriate configurations throughout the process. The temperature changes used in PCR essentially mirror these activities and corresponding configurations of DNA, in a more simplistic way. *See id.*

24. Fragkos et al., *supra* note 17.

Generally, primers sit approximately 1,000 bases apart.[25] PCR can produce millions of copies of DNA sequences by changing the kinetics of naturally occurring *in vivo* DNA replication into an exponential amplification process.

3. *Sanger Sequencing*

Like PCR, Sanger sequencing is another mimic of *in vivo* DNA replication. But, instead of exponentially amplifying DNA, Sanger sequencing determines the order of nucleotides in a DNA strand. The reagents required for Sanger sequencing are similar to those of PCR: (1) a single-stranded template DNA strand (to be sequenced); (2) nucleotides (free A, T, C, G to be added); (3) a sequencing primer (to establish a double-stranded foundation for sequencing to begin from); and (4) a polymerase enzyme (to catalyze the addition of each nucleotide to the growing complementary strand). Critically, Sanger sequencing reactions also include a fifth reagent: labeled dideoxynucleotides. There are two differences between labeled dideoxynucleotides and standard nucleotides, which, together, enable DNA sequencing.[26] Labeled dideoxynucleotides are: (1) modified to omit the 3'-OH group in the deoxyribose sugar group of their structure (hence the *dideoxy* prefix); and (2) tagged with a fluorescent dye, with each of A, T, C, and G having a different dye color (hence the *labeled* preface).

Sanger sequencing typically begins with PCR; having multiple copies of the template strand to be sequenced boosts efficiency.[27] Researchers extract the entire composite of genomic DNA from a target organism to use as a template, but then define the exact boundaries of a small template region using primer sequences. Once amplification of this template region occurs, Sanger sequencing begins on all PCR-amplified copies of this template at once. Again, the kinetics of Sanger sequencing reactions are the same as for *in vivo* DNA replication and PCR: the polymerase enzyme attaches to and physically moves along template DNA strands, sequentially adding nucleotides to a new strand based on complementarity to bases in the template strands, producing freshly synthesized, complementary DNA chains that mirror the templates.[28]

However, during Sanger sequencing, the polymerase enzyme occasionally adds a labeled dideoxynucleotide to a growing complementary DNA chain, instead of a standard, unmodified nucleotide. This happens randomly among all the growing DNA strands in the sequencing reaction—some strands start

25. *See* Pelt-Verkuil et al., *supra* note 21, at 11. At template lengths longer than 1,000 base pairs, fidelity and efficiency of the PCR process begin to decline.

26. For a graphic illustration of different nucleotide structures, see *infra* Figure 2.

27. Sigma on Sanger, *supra* note 16.

28. *Id.*

with a labeled dideoxynucleotide at the first possible position, some strands include one following many standard nucleotides, and some strands complete elongation entirely without ever adding one. Each time the polymerase adds a labeled dideoxynucleotide, the elongation of the growing DNA chain terminates at the position of incorporation. Chain termination occurs because the labeled dideoxynucleotides lack the 3'-OH required for addition of the next nucleotide in the DNA chain.[29] The labeled dideoxynucleotides also "color code" the terminated DNA chains with a unique fluorescent dye corresponding to the terminating nucleotide (**Figure 1C**). This process, importantly, is irreversible—the DNA chain cannot resume elongation once a labeled dideoxynucleotide has been added. Sanger sequencing is sometimes aptly called "chain-termination" sequencing.[30]

Therefore, like PCR, Sanger sequencing generates many copies of DNA, originating from a template strand that is typically no more than 1,000 base pairs in length.[31] However, all DNA copies generated by PCR are of the same length, mirroring the entire template sequence initially selected for amplification. Sanger-generated DNA copies are non-uniform in length because of the random processes of labeled dideoxynucleotide addition and subsequent chain termination. That is, after the sequencing reaction, the resulting product will include every possible length of DNA fragment, up to the full template length.[32] These fragments are referred to as oligonucleotides.

The different chain-terminated oligonucleotide lengths allow researchers to deduce the order of nucleotides in a template DNA strand. First, researchers will use gel electrophoresis to physically separate the chain-terminated oligonucleotides and arrange DNA fragments based on size. This process essentially lines up each chain-terminated oligonucleotide in order of decreasing size, from top to bottom on a gel.[33] Then, laser excitation of the fluorescent tags on each dideoxynucleotide enables researchers to visualize the physical distribution of the DNA fragments. Each DNA fragment shows up as a color-coded "band" on the gel, depending on the type of labeled dideoxynucleotide added to the final position on each fragment. Researchers can "read" these color-coded bands from smallest to largest, indicating the exact sequence of nucleotides from the first to last position of the template

29. The 3'-OH group participates in phosphodiester bond formation in typical strand elongation. *Id.*

30. *Id.*

31. *Id.* More typically, template fragments are 300–500 base pairs long.

32. To illustrate with an oversimplified example: a template strand that is 100 base pairs long will generate fragments of 100 different lengths: 1 base pair long, 2 base pairs long, 3 base pairs long, up until 100 base pairs long.

33. Sigma on Sanger, *supra* note 16.

DNA strand. This is a time-intensive process, as preparing gels for electrophoresis and running out DNA fragments is quite laborious.

The steps described in this Section illustrate the sequencing of just *one* template region of DNA, which practically cannot exceed approximately 1,000 bases.[34] To determine the entire genome sequence of an organism, Sanger sequencing must be repeated in 1,000 base pair increments. The haploid human genome is 3.055 billion base pairs long[35]—making Sanger sequencing prohibitively low-throughput for many modern applications.[36] However, Sanger sequencing remains the "gold-standard" for molecular biologists to sequence short regions of DNA (i.e., individual genes rather than entire genomes), with unmatched accuracy and fidelity compared to other techniques, including NGS.[37]

Despite its bottlenecked throughput, researchers used Sanger sequencing to sequence entire genomes before alternative approaches were developed.[38] In doing so, given the 1,000 base pair limitation of Sanger sequencing, researchers would have to process an entire genome into multiple 1,000 base pair regions, and then computationally stick them back together (formally termed "assembly") using a "shotgun" approach.[39] Some scientists initially preferred the idea of implementing a highly ordered process; that is, for a

---

34. The threshold of 1,000 base pairs is generally considered to be the maximum length of a template for Sanger sequencing. Beyond this, quality and accuracy plummet, as the size separation gel electrophoresis step of Sanger becomes unable to separate DNA fragments at an appropriate resolution. Henrik Stranneheim & Joakim Lundeberg, *Stepping Stones in DNA Sequencing*, 7 BIOTECHNOLOGY J. 1063 (2012).

35. Sergey Nurk et al., *The Complete Sequence of a Human Genome*, 376 SCI. 44 (2022).

36. One study estimated the reagents needed for Sanger sequencing to cost ~$500/Mb, and for NGS to cost $0.50/Mb. PHG FOUND., NEXT STEPS IN THE SEQUENCE: THE IMPLICATIONS OF WHOLE GENOME SEQUENCING FOR HEALTH IN THE UK 31 (2011), https://www.phgfoundation.org/media/140/download/Next%20steps%20in%20the%20sequence.pdf?v=1&inline=1.

37. *What is Next-Generation Sequencing (NGS)?*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/life-science/sequencing/sequencing-learning-center/next-generation-sequencing-information/ngs-basics/what-is-next-generation-sequencing.html (noting that NGS results are often verified using Sanger sequencing); *see also* Gagan & Van Allen, *supra* note 8, at 2 (addressing the loss in coverage (the depth of sequencing) and accuracy that occurs when the genic length to be sequenced is increased); *Key Differences Between Next-Generation Sequencing and Sanger Sequencing*, ILLUMINA, https://www.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html (last visited Oct. 2, 2022) (advertising that while NGS is more cost-effective for high numbers of gene targets, Sanger sequence is more cost-effective for low (e.g., 1-20) numbers of gene targets) [hereinafter Illumina on NGS vs Sanger].

38. *See* discussion *infra* Part III.

39. *See* Robert H. Waterston et al., *On the Sequencing of the Human Genome*, 99 PROCS. NAT'L ACAD. SCIS. 3712, 3712 (2002).

10,000 base pair genome, the first 1,000 base pair fragment might be from base 1 to base 1,000, the next fragment from base 500 to 1,500 (to maintain some overlap in case of inaccuracies at the tail ends), and so on. But such an ordered process required the assembler to have some form of mental "map" of the entire genome before beginning the process. The subsequent development of "shotgun" Sanger sequencing overcame this "map" requirement. With shotgun sequencing, researchers randomly break up ("shear") the genomic DNA into small fragments, sequence the fragments without a precise idea of their order, and then computationally assemble a genome sequence by comparing the base pairs that overlap between the fragments.[40] The sequencing products of the fragments are called *reads*; the reads after they have been assembled in the correct order are called *contigs*.[41]

40. Many were involved in the formulation of the shotgun strategy in the context of Sanger sequencing. A description of the shotgun approach closely followed the first articulation of Sanger sequencing in 1977. It seems that Rodger Staden was the first to suggest a shotgun strategy, in 1979. Rodger Staden, *A Strategy of DNA Sequencing Employing Computer Programs*, 6 NUCLEIC ACIDS RSCH. 2601 (1979). Then, Frederick Sanger published another report elaborating on the concept in 1980. Frederick Sanger et al., *Cloning in Single-Stranded Bacteriophage as an Aid to Rapid DNA Sequencing*, 143 J. MOLECULAR BIOLOGY 161 (1980). Joachim Messing followed similarly in 1981. Joachim Messing et al., *A System for Shotgun DNA Sequencing*, 9 NUCLEIC ACIDS RSCH. 309 (1981). Then, finally, Sanger applied the approach to a real genome sequence in 1982. Frederick Sanger et al., *Nucleotide Sequence of Bacteriophage λ DNA*, 162 J. MOLECULAR BIOLOGY 729 (1982).

41. Waterston et al., *supra* note 39, at 3712.

Figure 1: Foundational technology for DNA sequencing.

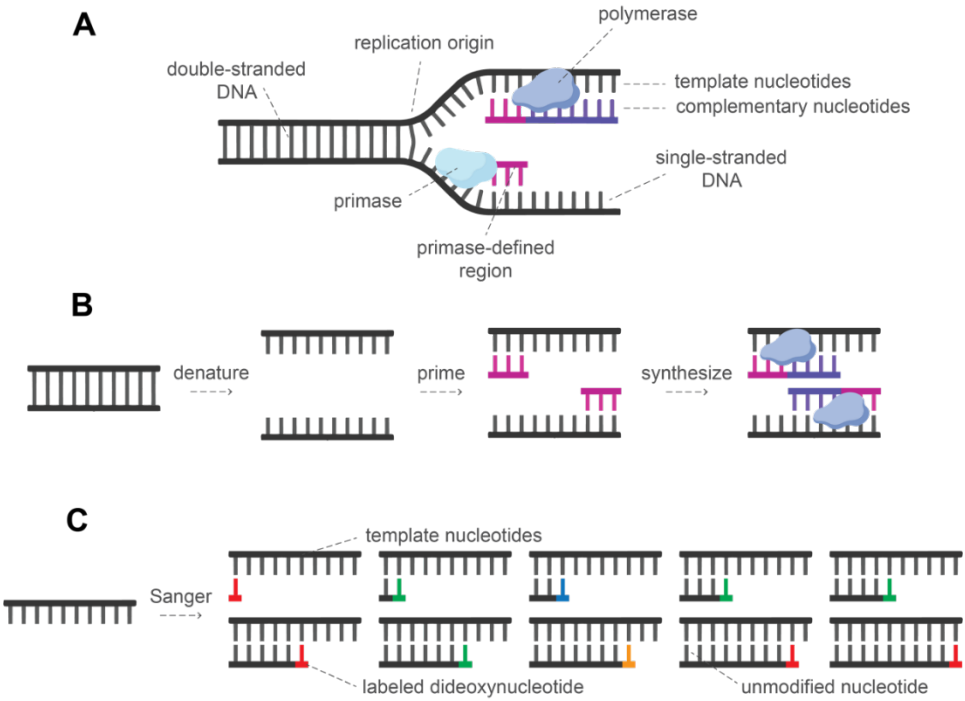


Figure 1 provides illustrations of the three foundational genetic replication processes described in Section II.A. **(A)** *In vivo* DNA replication requires the enzymatic activity of both polymerase (dark blue) and primase (light blue) enzymes (among others). After separation of double-stranded DNA into single-stranded DNA available for polymerase-mediated synthesis activity, primase enzymes introduce small defined regions from which replication can begin. Polymerase enzymes add complementary nucleotides to template single-stranded regions of DNA. **(B)** PCR mirrors the kinetic steps of in vivo DNA replication, beginning with initial denaturation of double-stranded DNA into single-stranded regions, priming with artificial oligonucleotides, and then using synthetic polymerase enzymes (dark blue) to grow a complementary DNA strand. **(C)** Sanger sequencing introduces unmodified nucleotides in iterative rounds of limited pseudo-replication, but occasionally adds labeled dideoxynucleotides for strand identification (red, green, blue, orange).

B. MODERN NGS TECHNOLOGY

Beginning in the early 2000s, the scientific community shifted away from Sanger sequencing toward NGS technologies.[42] NGS is a broad term that encompasses "massively parallel," high-throughput sequencing methods.[43] It is easiest to define NGS in reference to the previous "first generation" approach: Sanger sequencing examines individual template DNA strands in sequencing reactions that occur in separate environments; NGS examines millions of template DNA strands in parallel sequencing reactions, all in the same environment. With NGS, scientists can sequence the entire human genome in one day, for approximately $1,000.[44] Generally, in the context of DNA sequencing, researchers use NGS to sequence whole genomes, or large target regions within a genome, rather than individual genes.[45]

All NGS platforms apply the same basic approach, consisting of four steps.[46]

1. **Library preparation.** As in shotgun Sanger sequencing, researchers extract genomic DNA from a target organism, then randomly shear it into smaller fragments to use as template strands.
2. **Amplification.** Again, as in Sanger sequencing, PCR makes several copies of each fragmented template strand, to boost efficiency and generate a sufficient amount of substrate material.
3. **Read generation.** NGS platforms vary the most amongst each other and from Sanger sequencing in this step. Instead of the chain-terminating, dideoxynucleotide-based method of Sanger sequencing, some higher-throughput version of read generation occurs at this stage.[47]
4. **Data analysis.** Depending on the read generation method used in the third step, base calling and contig assembly proceed using various computational approaches. Briefly, researchers reassemble the entire

42. Michael L. Metzker, *Sequencing Technologies—The Next Generation*, 11 NATURE REVS. GENETICS 31, 31 (2010).

43. Dale Muzzey et al., *Understanding the Basics of NGS: From Mechanism to Variant Calling*, 3 CURRENT GENETIC MED. REPS. 158, 159 (2015) (defining NGS as "a diverse collection of post-Sanger sequencing technologies").

44. *Id.* at 158–59.

45. Illumina on NGS vs Sanger, *supra* note 37.

46. *See* Keegan Schroeder, *A History of Sequencing*, FRONTLINE GENOMICS (Apr. 19, 2022), https://frontlinegenomics.com/a-history-of-sequencing/.

47. For a description of other non-Illumina NGS platforms and the read generation techniques used at this stage, see discussion *infra* Part V.

genome sequence from the short shotgun-fragmented strands of DNA.[48]

Although there are countless NGS platforms and read generation approaches, Illumina sequencing technology—the focus of this Article—dominates the NGS market. In late 2019, the Federal Trade Commission (FTC) characterized Illumina as allegedly generating more than 90% of the world's sequencing data.[49] The characteristic, "massively parallel" aspect of Illumina sequencing arises from three unique elements of Illumina, described in the next Section.

1. Three Illumina Elements

The three unique elements of Illumina are integrated into steps two (amplification) and three (read generation) of the sequencing pipeline: (1) the use of a solid support (step two); (2) the bridge PCR amplification of DNA fragments to generate clusters (step two); and (3) the technique of sequencing-by-synthesis (step three). This Section describes each element.

a) Solid Support Array

The first "massively parallel" element of Illumina technology is a solid support to physically attach template DNA strands prior to PCR amplification, in contrast to the aqueous suspension of DNA fragments in Sanger sequencing.[50] Illumina uses a solid support called a flow cell, which is coated with a lawn of two types of oligonucleotides (short DNA strands) physically anchored to the flow cell surface.[51] After breaking the genomic DNA of the target organism into smaller fragments in step one (library preparation), researchers attach two types of adapters to the ends of each single-stranded fragment through a process called ligation.[52] All the fragments have the same type of adapter at their "start" (called the 5' end) and a different type of adapter at their "end" (called the 3' end).[53] Both the 5' adapter and the 3' adapter are complementary to the two types of oligonucleotides anchored to the flow cell surface, such that the entire library of template DNA strands bind to the

48. *A Brief Guide to Genomics*, *supra* note 3.

49. Complaint ¶¶ 1, 34, 35, Illumina, Inc. & Pacific Biosciences California, Inc. v. F.T.C., No. 9387 (Dec. 17, 2019).

50. *See* Muzzey et al., *supra* note 43, at 159.

51. *More Data, Reduced Costs, and Faster Runs*, ILLUMINA, https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/patterned-flow-cells.html (last visited Nov. 24, 2022).

52. *Uniformity, Precision and Reliability in Library Preparation*, ILLUMINA, https://www.illumina.com/techniques/sequencing/ngs-library-prep/ligation.html (last visited Nov. 24, 2022).

53. *Id.*

oligonucleotides, at either their 5' or 3' ends.[54] This process physically anchors all the template strands to the flow cell in a random array.[55] While Sanger sequencing maintains DNA fragments in suspension and carries out *size separation* after the sequencing reactions are completed, Illumina sequencing uses the solid support to establish *positional separation* between DNA fragments before the sequencing reactions begin.

b) Bridge PCR Clustering

The second "massively parallel" element of Illumina technology is bridge PCR clustering to amplify template DNA strands, instead of the standard PCR step conducted prior to Sanger sequencing. This bridge PCR process requires that DNA be fixed to a solid support, as in the Illumina platform. In routine PCR, as described in Section II.A.2 *supra*, polymerase enzymes repeatedly synthesize complementary strands of DNA from template strands, producing double-stranded DNA fragments that are repeatedly denatured for iterative rounds of amplification. This process occurs stochastically in liquid suspension. For the PCR that occurs during Illumina sequencing, some fraction of template strands must always remain physically anchored to the flow cell throughout the amplification process, complicating the requirement for repeated denaturation and iterative amplification.

The Illumina platform solves this anchored denaturing problem with 5' and 3' adapter sequences and complementary oligonucleotides on the flow cell. After each template strand attaches to the flow cell at one end, the strands fold over and form a bridge with the oligonucleotide complementary to the adapter sequence at the other end.[56] That is, a template strand bound to the flow cell at its 5' end folds over and binds to a different oligonucleotide, complementary to its 3' end.[57] After this, the kinetics of bridge PCR follows routine PCR, with similar reagents: (1) a single-stranded template DNA strand, in bridge format (to be amplified); (2) nucleotides (free A, T, C, G to be added); and (3) a polymerase enzyme (to catalyze the addition of each nucleotide to the growing strand copy).[58] A polymerase enzyme attaches to the adapter-oligonucleotide

54. *Id.*

55. ILLUMINA, AN INTRODUCTION TO NEXT-GENERATION SEQUENCING FOR CARDIOLOGY 4 (2015) [hereinafter ILLUMINA GUIDE].

56. *Id.* at 3, 7 (2015).

57. James M. Heather & Benjamin Chain, *The Sequence of Sequencers: The History of Sequencing DNA*, 107 GENOMICS 1, 3 (2016).

58. Unlike routine PCR, forward and reverse primers are not needed for bridge PCR clustering, as the binding of the template DNA strand adapters to the flow cell-anchored oligonucleotides creates the double-stranded foundations that polymerase enzymes require for attachment.

paired region and physically moves along the template DNA strand bridge, sequentially adding nucleotides to a new strand based on complementarity to bases in the template. The resulting product is a double-stranded bridge, rather than the linearized double-stranded DNA chain of routine PCR. The bridge then denatures in response to the same temperature trigger as in routine PCR, so the original template strand and the newly synthesized complementary strand release from the flow cell at one end and remain anchored physically to the flow cell at only the 5' or 3' end.[59]

The bridging, amplification, and denaturation process repeats itself iteratively, for every unique template strand fragment distributed randomly throughout the flow cell. Importantly, Illumina sequencing platforms have a maximum read length of 300 base pairs, with 150 base pair reads as the most common length.[60] This short length—even shorter than the read length used in Sanger sequencing—means that each template strand folds over and forms a bridge more frequently with complementary oligonucleotides that are physically proximal to the original oligonucleotide anchor. Thus, bridge PCR produces a characteristic *clustering* effect, as the bridges continue to form in the same, localized area, outwards from each template strand fragment.[61] In other words, the resulting DNA lawn preserves the positioning of the initial fragments of unique template DNA strands, with entire clusters of template DNA strands positionally separated.

c) Sequencing-by-Synthesis (SBS) Read Generation

The third "massively parallel" element of Illumina technology is SBS, which replaces the chain termination aspect of Sanger sequencing. Among the three critical elements outlined in this Section, Illumina's unique approach to SBS is the most essential component of its platform.[62] As in Sanger sequencing reactions, SBS reactions also use: (1) a single-stranded template DNA strand; (2) a sequencing primer; and (3) a polymerase enzyme. However, rather than standard, unmodified nucleotides or labeled dideoxynucleotides, SBS reactions

59. *See* ILLUMINA GUIDE, *supra* note 55, at 4.

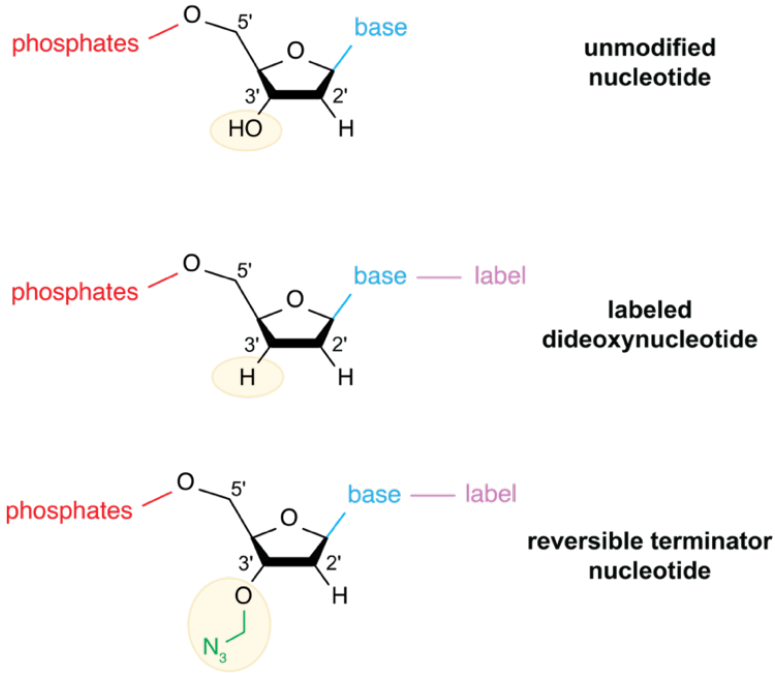
60. *Maximum Read Length for Illumina Sequencing Platforms*, ILLUMINA, https://support.illumina.com/bulletins/2020/04/maximum-read-length-for-illumina-sequencing-platforms.html (last visited Oct. 2, 2022). Illumina sequencing is limited to relatively short read lengths to preserve accuracy—at lengths longer than 300 base pairs, fidelity of the sequencing process declines. *Id.*

61. *See* ILLUMINA GUIDE, *supra* note 55, at 3–4.

62. *See* Jason A. Reuter et al., *High-Throughput Sequencing Technologies*, 58 MOLECULAR CELL 586, 586 (2015) (summarizing developments in NGS technologies).

use only a single nucleotide type: reversible terminator nucleotides (**Figure 2**).[63]

Figure 2: SBS chemistry in the Illumina NGS platform.



There are two critical differences between reversible terminator nucleotides and standard nucleotides. Reversible terminator nucleotides: (1) have a 3'-*O*-blocking group instead of the 3'-OH group in the deoxyribose sugar group; and (2) are tagged with a cleavable fluorescent dye, with each of A, T, C, and G having a different dye color.[64] These nucleotides may seem similar to the labeled dideoxynucleotides of Sanger—but they have unique chemistry. While labeled dideoxynucleotides completely lack the 3'-OH group in the deoxyribose sugar, reversible terminator nucleotides simply have a blocking group added to the 3'-OH.[65] Uniquely, this blocking group can be chemically cleaved off. And while both labeled dideoxynucleotides and reversible terminator nucleotides are tagged with a fluorescent dye, the tag on reversible terminator nucleotides—just like the blocking group—can be

63. ILLUMINA GUIDE, *supra* note 55, at 3.

64. David R. Bentley et al., *Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry*, 456 NATURE 53, 53 (2008) (sequencing a human genome).

65. *Id.*

chemically cleaved off.[66] Critically, the addition of a *single* chemical reagent can simultaneously cleave both the 3'-*O*-blocking group and the fluorescent tag.[67]

The SBS process includes several steps in common with *in vivo* DNA replication, PCR, and Sanger sequencing: the polymerase enzyme attaches to and physically moves along template DNA strands, sequentially adding nucleotides to a new strand based on complementarity to bases in the templates and producing freshly synthesized chains of DNA complementary to the templates. The difference in SBS, compared to Sanger sequencing, is that *only* reversible terminator nucleotides are incorporated into growing complementary DNA chains, not standard nucleotides or labeled dideoxynucleotides. And instead of the *irreversible* chain termination that stochastically occurs with the addition of a labeled dideoxynucleotide to growing DNA strands in Sanger sequencing, the addition of a reversible terminator nucleotide results in *reversible* chain termination of growing DNA strands in SBS. In Sanger sequencing, labeled dideoxynucleotides irreversibly terminate chain elongation because they lack the 3'-OH group of standard nucleotides. On the other hand, in SBS, reversible terminator nucleotides reversibly pause chain elongation because of the 3'-*O*-blocking group.[68] After a chemical reagent is added to cleave off the blocking group, chain elongation resumes with the addition of the next reversible terminator nucleotide.[69] Cleavage of the blocking group also removes the fluorescent dye so that a new color code can be introduced with the next nucleotide.[70]

This mechanism of reversible termination separates SBS from Sanger sequencing in two ways. First, Sanger sequencing reactions generate entire libraries of oligonucleotides of varying lengths, with each one permanently color-coded based on the terminal labeled dideoxynucleotide. SBS reactions only generate oligonucleotides of the same length as the template strand, and color-coding exists only transiently, between the moment of incorporation of a reversible terminator nucleotide and the subsequent cleavage of its

66. *Sequencing-by-Synthesis: Explaining the Illumina Sequencing Technology*, BITESIZEBIO (Aug. 30, 2012), https://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/. In the Illumina reversible terminator nucleotides, the fluorescent tag is attached to the nucleobase.

67. *Id.* At the 3' position, the reagent removes the blocking group and regenerates a 3'-OH group so strand elongation can proceed; at the position where the fluorescent dye is attached, the dye itself is removed but a scar remains in its place. *Id.*

68. Bentley et al., *supra* note 64, at 53.

69. *Explore Illumina Sequencing Technology*, ILLUMINA, https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html (last visited Sept. 27, 2022). As in Sanger sequencing at the gel imaging step, a characteristic fluorescent signal is emitted per nucleotide type (A, T, C, G).

70. *See* Bentley et al., *supra* note 64, at 53.

fluorescent tag. Second, the products of Sanger sequencing reactions must be separated based on size through gel electrophoresis, to sort through the multi-length oligonucleotide library generated from the process and to visualize the fluorescent labels through laser excitation. SBS replaces this labor-intensive separation process with laser excitation integrated directly into the sequencing platform.

At the exact moment that a reversible terminator nucleotide incorporates into a growing DNA chain, elongation pauses due to the 3'-*O*-blocking group, a laser excites the fluorescent tag, and the system records the emitted signal for the corresponding spatial position in the template strand. This process is called "base calling." Then, after cleavage of both the 3'-*O*-blocking group and the fluorescent tag, the next reversible terminator nucleotide incorporates, elongation pauses, and a laser excites the new fluorescent tag corresponding to the just-added nucleotide once again, at a spatial position one "layer" above the previous signal. Base calling, or sequencing, occurs continuously, in real-time, as strands elongate—hence, the name sequencing-*by-synthesis*.

2. *Integration into "Massively Parallel" Sequencing*

Together, the three elements outlined for SBS, *supra*, allow for synchronous sequencing of millions of DNA fragments in real-time. First, positional separation produced by the anchoring of fragments to a solid support locks each fragment in a single position. Second, bridge PCR and cluster generation maintain this positional separation through the amplification step and subsequent sequencing reactions, providing an adequate signal for base calling. Third, the use of reversible terminator chemistry in SBS transiently color-codes each incorporated nucleotide, so that base calling may occur at the moment of nucleotide incorporation into each template strand during continuous growth in a fixed cluster. The combination of these elements enables laser excitation to image a bird's eye view of the entire flow cell. After the incorporation of each reversible terminator nucleotide to each template strand, the system captures an image that depicts, for each cluster, the fluorescent signal from the last nucleotide added to the growing DNA chains.[71] Therefore, millions of strands complementary to the template strand are read simultaneously within each cluster, and millions of clusters are sequenced simultaneously within the flow cell.

C. LIFE SCIENCES APPLICATIONS OF NGS

DNA sequencing can link observable, health-relevant phenotypes (i.e., observable physical characteristics) with underlying genotypes (i.e., DNA

71. ILLUMINA GUIDE, *supra* note 55, at 3–4.

sequences). DNA contains the instructional material for the synthesis of mRNA, which, in turn, contains the instructional material for protein synthesis.[72] Proteins control our biological functions.[73] Mutated DNA can create aberrant proteins, which can produce disease-causing abnormalities—"[v]irtually every human ailment has some basis in our genes."[74]

Because NGS can sequence whole human genomes at a much faster rate than Sanger sequencing, NGS empowers scientists to identify many potentially causal genetic differences ("variants") between patient genomes and healthy ("reference") genomes.[75] NGS has radically improved three main areas of life sciences applications: diagnostic testing, personalized medicine, and direct-to-consumer genomics.

1. *Diagnostic Testing*

NGS revolutionized the diagnostics field and is now a routine and increasingly affordable technique to identify disease-indicating genetic variants in patient DNA samples. Unlike Sanger sequencing, NGS users need not know what type of genetic variant they are looking for, nor where in a patient's genome to look for it.[76] In complex diseases with multiple underlying mutations in coding and non-coding regions of DNA—and often a complete lack of prior knowledge—this is a critical advantage.[77] For example, doctors previously diagnosed many subtypes of cancer based on morphology or other phenotypic signatures; now, they can distinguish cancers from genetic profiles at earlier stages[78]—"an unattainable fantasy" prior to the advent of NGS.[79] More generally, NGS also facilitates genome-wide association studies that correlate variants to disease phenotypes at the population level using statistical analyses.[80] These large-scale studies generate pools of data that help optimize

72. *A Brief Guide to Genomics*, *supra* note 3.

73. *Id.*

74. *Id.*

75. Muzzey et al., *supra* note 43, at 158 (dividing variants of interest into (1) changes to DNA sequences (e.g., single nucleotide polymorphisms, insertions, deletions), and (2) large deletions or duplications of whole genes). A genome—the target molecule of genomics—is the entire composite of DNA within an organism, stored in linearized or circular form.

76. *See* Sam Behjati & Patrick S. Tarpey, *What is Next Generation Sequencing?*, 98 ARCHIVES DISEASE CHILDHOOD 236, 236 (2013).

77. *Complex disease genomics*, ILLUMINA, https://www.illumina.com/areas-of-interest/complex-disease-genomics.html (last visited Sept 23, 2022).

78. *See* Gagan & Van Allen, *supra* note 8, at 5.

79. Stratton et al., *The Cancer Genome*, 458 NATURE 719, 722–23 (2009) (stating that "the arrival of second-generation sequencing technologies promise[d] a new era for cancer genomics").

80. *Disease association studies*, ILLUMINA, https://www.illumina.com/areas-of-interest/complex-disease-genomics/gwas.html (last visited Sept. 23, 2022).

how physicians approach disease screening to guide more targeted diagnostic approaches.[81]

2. Personalized Medicine

When NGS identifies variants for disease diagnostics, the variants themselves might be the root cause of the disease of interest. Other times, those variants might simply be associated with the presence of disease for an unknown reason. But non-causal "associated" variants might still indicate something useful for personalizing disease treatment.[82] For example, certain genetic mutations increase the likelihood that a patient will either respond to or resist a therapeutic strategy. With a patient's genetic profile, a physician might be able to select a specific type of chemotherapy or treatment approach.[83] And as sequencing methods continue to improve in both speed and miniaturization, physicians can make personalized decisions based on genetic information at or close to the point-of-care using portable technologies, even for rare or novel genetic mutations.[84] Integrating genotypic assessments into clinical examinations means physicians can consider genetic data holistically along with pathology and symptom assessments.[85]

3. Direct-to-Consumer Genomics

The efficiency of NGS technologies has made it possible to sell personalized genetic testing kits to interested consumers, allowing for general

---

81. *See id.* Genome-wide association studies provide correlational evidence of variants that are present at different frequencies in human populations lacking a certain disease, compared to healthy populations. David J. Hunter et al., *Letting the Genome Out of the Bottle – Will We Get Our Wish?*, 358 NEW ENG. J. MED. 105, 105 (2008).

82. *See* Gagan & Van Allen, *supra* note 8, at 8 ("NGS is inextricably intertwined with the realization of precision medicine in oncology.").

83. *See* Monica Avila & Funda Meric-Bernstam, *Next-Generation Sequencing for the General Cancer Patient*, 17 CLINICAL ADVANCES HEMATOLOGY & ONCOLOGY 447 (2019); Gagan & Van Allen, *supra* note 8, at 6 (listing several types of disease for which certain DNA mutations are indications or contraindications for therapeutic approaches in Table 2).

84. *See, e.g.*, Brandon S. Sheffield et al., *Point of Care Molecular Testing: Community-Based Rapid Next-Generation Sequencing to Support Cancer Care*, 29 CURRENT ONCOLOGY 1326 (2022) (discussing one example of NGS use in a clinical setting, where a workflow was implemented to get genetic profiling results back to patients in 3 business days).

85. *See, e.g.*, Yaoting Gui et al., *Frequent Mutations of Chromatin Remodeling Genes in Transitional Cell Carcinoma of the Bladder*, 43 NATURE GENETICS 875 (2011) (bladder); Guangwu Guo et al., *Frequent Mutations of Genes Encoding Ubiquitin-Mediated Proteolysis Pathway Components in Clear Cell Renal Cell Carcinoma*, 44 NATURE GENETICS 17 (2011) (kidney); Michael F. Berger et al., *The Genomic Complexity of Primary Human Prostate Cancer*, 470 NATURE 214 (2011) (prostate); Xose S. Puente et al., *Whole-Genome Sequencing Identifies Recurrent Mutations in Chronic Lymphocytic Leukemia* 475 NATURE 101 (2012) (CLL); Timothy J. Ley et al., *DNA Sequencing of a Cytogenetically Normal Acute Myeloid Leukemia Genome*, 456 NATURE 66 (2008) (AML).

susceptibility testing and genetic profiling.[86] Genetic testing irrespective of disease state can facilitate early surveillance and detection in some populations, if interpreted properly and paired with appropriate medical direction.[87]

III. DEVELOPMENT OF THE ILLUMINA NGS PLATFORM

The Illumina NGS platform has a long history, from early developments in first-generation sequencing in the 1970s, to the massive technological leap pushed forward by the Solexa scientists at the turn of the century. This Part chronicles this history in five phases: (1) optimization and commercialization of Sanger sequencing; (2) implementation of Sanger sequencing in the HGP; (3) preliminary research driving key pre-Illumina advances in NGS; (4) creation of the NGS Solexa idea; and (5) expansion and commercialization of Solexa (now Illumina).

A. PHASE 1: OPTIMIZATION AND COMMERCIALIZATION OF SANGER SEQUENCING

Frederick Sanger, a biochemist at the Laboratory of Molecular Biology funded through the Medical Research Council in the United Kingdom, published a description of the first form of "Sanger sequencing" in 1977.[88] At approximately the same time, Harvard scientists Allan Maxam and Walter Gilbert independently developed a similar approach.[89] Their method—termed Maxam-Gilbert sequencing—was initially more popular, but fell out of favor as scientists recognized the comparative technical ease of Sanger sequencing.[90]

---

86. *See* Hunter et al., *supra* note 81.

87. *See id.* (warning that such test kits may be inaccurate and yield false positives, and that consumers may incorrectly interpret results without appropriate guidance). Some have termed at-home genetic testing "recreational genomics," and remarked that this phenomenon carries high risks of misinformation. James P. Evans, *Recreational Genomics; What's in it for You?*, 10 GENETICS MED. 709, 710 (2008).

88. Frederick Sanger et al., *DNA Sequencing with Chain-Terminating Inhibitors*, 74 PROCS. NAT'L ACAD. SCIS. 5463 (1977). Sanger's 1977 publication, cited here, is the first report using *dideoxynucleotides* in the sequencing reactions, producing the chain-terminating element of Sanger sequencing. However, it is worth noting that Sanger first published the "plus and minus" sequencing method in 1975, which was later refined in his 1977 publication. Frederick Sanger & Alan R. Coulson, *A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase*, 94 J. MOLECULAR BIOLOGY 441 (1975).

89. Maxam-Gilbert sequencing was first reported in 1977. This method still uses chain-termination, but not due to an intrinsic structural modification of the nucleotides to be incorporated (i.e., no *dideoxy* element). Allan M. Maxam & Walter Gilbert, *A New Method for Sequencing DNA*, 74 PROCS. NAT'L ACAD. SCIS. 560, 560 (1977).

90. *See* Christopher M. Holman, *Advances in DNA Sequencing Lead to Patent Disputes*, 30 NATURE BIOTECHNOLOGY 1054, 1054 (2012).

Immediately after the initial publication of Sanger sequencing, researchers began trying to automate the method, which set the stage for the eventual leap into NGS technologies. In its first form, Sanger sequencing involved an entirely manual process—for example, the process initially used radioactive labeling, rather than fluorescence.[91] Because all four nucleotides had the same type of tag (instead of four distinct color codes), users had to run four separate sequencing reactions for each DNA fragment to be sequenced, to track each nucleotide.[92] Then, in the final gel electrophoresis step, visualization had to occur via autoradiography rather than laser excitation and without computational automation of the base calling step.[93] Manual Sanger sequencing also used rectangular slab gels, which each require their own separate dock and typically run on a vertical or horizontal plane.[94]

Leroy Hood, a professor at the California Institute of Technology ("Caltech"), spearheaded the automation of Sanger sequencing. Beginning in the 1980s, Hood suggested fluorescence instead of radiolabeling, so the four nucleotides could each have their own color codes.[95] Fluorescent labelling enabled researchers to combine the four separate sequencing reactions into one and to image the fluorescent tags with simple laser excitation, rather than the lengthy autoradiography process.[96] James Prober then refined this labeling method, labeling the dideoxynucleotides themselves with fluorescent dyes, instead of the indirect primer-mediated tagging of Sanger and Hood's preliminary methods.[97] Other scientists also proposed replacing the manual gel electrophoresis step with capillary electrophoresis.[98] Instead of rectangular gel slabs, capillary electrophoresis uses gels polymerized in capillary tubes, arrayed

91. Jeffrey M. Perkel, *An Automated DNA Sequencer*, 18 SCIENTIST 40, 40 (2004).

92. *Id.*; Sanger, *supra* note 88, at 5464.

93. Sanger, *supra* note 88, at 5464.

94. *See id.*

95. Lloyd M. Smith et al., *The Synthesis of Oligonucleotides Containing an Aliphatic Amino Group at the 5' Terminus: Synthesis of Fluorescent DNA Primers for Use in DNA Sequence Analysis*, 13 NUCLEIC ACIDS RSCH. 2399 (1985).

96. Lloyd M. Smith et al., *Fluorescence Detection in Automated DNA Sequence Analysis*, 321 NATURE 674 (1986); *see* Schroeder, *supra* note 46. The fluorescent readout of the laser excitation is then computationally processed to generate chromatograms. These are four-color plots that depict color-coded nucleotide "peaks," corresponding to the fluorescent signals emitted from each type of nucleotide. Researchers examine chromatograms to infer the identity and order of the base pairs in a sequenced DNA strand.

97. James M. Prober et al., *A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides*, 238 SCI. 336 (1987).

98. Aharon S. Cohen et al., *Rapid Separation and Purification of Oligonucleotides by High-Performance Capillary Gel Electrophoresis*, 85 PROCS. NAT'L ACAD. SCIS. 9660 (1988); John A. Luckey et al., *High Speed DNA Sequencing by Capillary Electrophoresis*, 18 NUCLEIC ACIDS RSCH. 4417 (1990).

in 384-tube format.[99] Improvements to the labeling and the electrophoresis processes combined with an integrated laser detection system allowed for "automated" Sanger sequencing. In 1985, Hood developed the first machine implementing automated Sanger Sequencing (the ABI 370) at his then-newly founded company, Applied Biosystems.[100]

B. PHASE 2: IMPLEMENTATION OF SANGER SEQUENCING IN THE HUMAN GENOME PROJECT

With the inclusion of fluorescent dyes, capillary electrophoresis, and automated laser detection, DNA sequencing exploded in popularity.[101] In addition to sequencing the genomes of several small organisms,[102] the scientific community sought to use Sanger sequencing to determine the complete sequence of the human genome.[103] This idea, at first, was polarizing. Some felt that the genome contained mostly "junk" DNA[104] and its sequence would be a useless resource, and that focusing on such "big" science would "divert[]

99. *Electrophoresis with Sanger Sequencing*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/life-science/sequencing/sanger-sequencing/sanger-dna-sequencing/electrophoresis-sanger-sequencing.html?socid=social_btb_abseq (last visited Mar. 14, 2024).

100. Marina Barba et al., *Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology*, 6 VIRUSES 106 (2014); *Leroy Hood*, LEMELSON-MIT, https://lemelson.mit.edu/award-winners/leroy-hood#:~:text=Working%20with%20a%20team%20at,strings%20of%20DNA%20in%20cells. (last visited May 2, 2023).

101. *See* Alice Maria Giani et al., *Long Walk to Genomics: History and Current Approaches to Genome Sequencing and Assembly*, 18 COMPUTATIONAL & STRUCTURAL BIOTECHNOLOGY J. 9, 11 (2020); Frederick Sanger, *Sequences, Sequences, and Sequences*, 57 ANN. REV. BIOCHEMISTRY 1, 25 (1988).

102. A complete genome sequence of a live organism (not bacteriophage) was reported for the first time in 1995. Robert D. Fleischmann et al., *Whole-Genome Random Sequencing and Assembly of Haemophilus influenzae Rd*, 269 SCI. 496 (1995) (reporting the use of Sanger sequencing to obtain the complete sequence of the ~ 1.8 million base pair genome of a *Haemophilus* bacterium, out of Craig Venter's group). Before this, a few groups had reported bacteriophage genome sequences. *See, e.g.*, Sanger, *supra* note 40.

103. *See* Reuter et al., *supra* note 62, at 586.

104. With an entire research field now devoted to the analysis of non-coding DNA, the "junk" DNA terminology has been more or less debunked. Indeed, several regions of the human genome do not encode for specific proteins. But those regions are often functionally critical for other purposes (e.g., to regulate gene expression). *The Complex Truth About 'Junk DNA'*, QUANTAMAGAZINE, https://www.quantamagazine.org/the-complex-truth-about-junk-dna-20210901/ (last visited Nov. 24, 2022); David Brown & Hristio Boytchev, *'Junk DNA' Concept Debunked by New Analysis of Human Genome*, WASH. POST (Sept. 5, 2012), https://www.washingtonpost.com/national/health-science/junk-dna-concept-debunked-by-new-analysis-of-human-genome/2012/09/05/cf296720-f772-11e1-8398-0327ab83ab91_story.html. Unfortunately, the "junk" terminology continues to plague the legal field, especially in the context of forensic analysis. Jennifer K. Wagner, *Out with the "Junk DNA" Phrase*, J. FORENSIC SCI. (2012).

resources from the 'real' small science."[105] Others felt that uncovering the genome sequence would establish "an unparalleled medical and research tool for studying mutations," "the grail of human genetics,"[106] and be "crucial for progress in human physiology and pathology," especially in the context of cancer research.[107] After several preliminary meetings and discussions in the late 1980s, the National Institutes of Health (NIH) and the Department of Energy formally initiated the HGP in October 1990—motivated primarily by a desire to understand the broad effects of radiation exposure on the human genome.[108] Formally, the HGP was an international agreement between researchers to work together on the production of a single reference human genome sequence.[109] Hood's improvements to Sanger sequencing—and the resulting automated sequencer machines that he developed—are credited as the technology that made the HGP possible.[110]

The HGP is now fondly regarded as "the largest single undertaking in the history of biological science,"[111] yielding sequencing data for "over 90% of the human genome."[112] However, the story of the HGP also illustrates the inability for Sanger sequencing, even automated, to keep pace with modern life sciences sequencing inquiries. The Project—which sequenced *one* human genome—cost an estimated $3 billion[113] and lasted for twelve years, ending only in April 2003.[114] With the lofty goal of sequencing *billions* of human genomes,

105. Leroy Hood & Lee Rowen, *The Human Genome Project: Big Science Transforms Biology and Medicine*, 5 GENOME MED. 1, 1 (2013).

106. Robert Kanigel, *The Genome Project*, N.Y. TIMES (Dec. 13, 1987), https://www.nytimes.com/1987/12/13/magazine/the-genome-project.html.

107. Renato Dulbecco, *A Turning Point in Cancer Research: Sequencing the Human Genome*, 231 SCI. 1055, 1055 (1986).

108. Hood & Rowen, *supra* note 105, at 1.

109. U.S. DEP'T HEALTH & HUM. SERVS. & U.S. DEP'T ENERGY, UNDERSTANDING OUR GENETIC INHERITANCE, THE U.S. HUMAN GENOME PROJECT: THE FIRST FIVE YEARS (1990).

110. Andrew Pollack, *SCIENTIST AT WORK: LEROY HOOD; A Biotech Superstar Looks at the Bigger Picture*, N.Y. TIMES (Apr. 17, 2001), https://www.nytimes.com/2001/04/17/science/scientist-at-work-leroy-hood-a-biotech-superstar-looks-at-the-bigger-picture.html.

111. SIMON TRIPP & MARTIN GRUEBER, ECONOMIC IMPACT OF THE HUMAN GENOME PROJECT (2011).

112. *Human Genome Project Fact Sheet*, NAT'L HUM. GENOME RSCH. INST., https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project (last visited Mar. 14, 2024) [hereinafter HGP Fact Sheet].

113. This estimate is based on the initially projected cost for the HGP, as "precise cost-accounting [is] difficult to carry out, especially across the set of international funders." *Id.*

114. *The Human Genome Project*, NAT'L HUM. GENOME RSCH. INST., https://www.genome.gov/human-genome-project (last visited Mar. 14, 2024) [hereinafter HGP Basics]. A draft of the human genome was initially published in 2001. Int'l Human Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 NATURE 860

researchers needed to improve sequencing technology throughput significantly.[115] The HGP largely motivated the development of NGS technologies and systems biology, which together precipitated the "omics" era.[116]

Some consider the HGP key to advancing the concepts of "open science" and data sharing, because such a complicated international effort required coordination between the participating researchers.[117] The HGP researchers (later termed the International Human Genome Sequencing Consortium) worked without commercial funding sources.[118] To effectively coordinate, the Consortium agreed on the "Bermuda Principles" in 1996, which required all participants to make their sequence data available in public databases within approximately twenty-four hours of generation.[119]

Other perspectives on the commercialization of the human genome loomed in the background. In the middle of 1998, with the HGP in full swing, Craig Venter announced his plans to found Celera Genomics and launch a competing effort to sequence the human genome using a variation of shotgun assembly that others dismissed as too computationally intensive for the human genome.[120] Venter's announcement—and his subsequent suggestion that the HGP should move on and try the mouse genome instead—sparked panic that

---

(2001) [hereinafter HGP First Draft]. The full sequence was finalized in a subsequent publication in 2004. Int'l Human Genome Sequencing Consortium, *Finishing the Euchromatic Sequence of the Human Genome*, 431 NATURE 931 (2004).

115. HGP Basics, *supra* note 114.

116. Muzzey et al., *supra* note 43, at 158 (noting that "by the end of the Human Genome Project in 2002, [Sanger sequencing] was already operating at nearly peak efficiency"); Hood & Rowen, *supra* note 105, at 5.

117. *See* Kendall Powell, *The Broken Promise that Undermines Human Genome Research*, NATURE NEWS FEATURE (Feb. 10, 2021), https://www.nature.com/articles/d41586-021-00331-5 (referring to David Haussler, one of the developers of the first web-based tool for viewing the human genome sequence, stating that before the HGP, "there had not been a serious discussion about data sharing in biomedical research," and that "[t]he standard was that a successful investigator held onto their own data as long as they could").

118. HGP Fact Sheet, *supra* note 112 (noting that the project was "one of the most ambitious and important scientific endeavors in human history," seeking to sequence the entire human genome and the genomes of several model organisms: bacteria, yeast, flies, nematodes, and mice). Funding for the HGP was congressionally approved through the National Institutes of Health and the Department of Energy, and also separately from the Wellcome Trust and the Medical Research Council in the United Kingdom.

119. *Id.* (noting that the Bermuda Principles are "credited with establishing a greater awareness and openness to the sharing of data in biomedical research," and that they are "one of the most important legacies of the [HGP]").

120. Jan Witkowski, *A Life Worth Writing About*, 449 NATURE 785, 786 (2007) (providing more information on Venter's storied career, as a scientist and entrepreneur); Waterston et al., *supra* note 39, at 3712.

a private company would own the human genome sequence and that Congress would pull funding from the HGP.[121] Venter indicated that Celera would seek patent protection on all gene sequences obtained from their effort, and would not comply with the Bermuda Principles.[122]

Galvanized by Venter's threat, the HGP continued in full force.[123] And Celera began a parallel sequencing effort shortly after Venter's announcement. In 2000, former U.S. President Bill Clinton and former U.K. Prime Minister Tony Blair delivered a joint statement advocating that the human genome sequence "be made freely available to scientists everywhere."[124] Shortly after, in 2001, both the HGP and Celera teams published their first drafts of the human genome, one day apart.[125] The HGP used Sanger sequencing with hierarchical "clone-by-clone" assembly; Celera used Sanger sequencing with whole-genome shotgun assembly to generate almost the exact same sequencing product as the HGP in one-tenth of the time.[126]

The HGP is now viewed as "instrumental in pushing the development of high-throughput [sequencing] technologies."[127] The frustratingly slow speed of the Project—exacerbated by the competitive environment sparked by the Celera effort—encouraged scientists to improve sequencing technologies.[128]

---

121. Waterston et al., *supra* note 39, at 3712.

122. Caroline Barranco, *The Human Genome Project*, NATURE MILESTONES (Feb. 10, 2021), https://www.nature.com/articles/d42859-020-00101-9 (last visited Oct. 1, 2022).

123. *See, e.g.*, *id.*; Hunter et al., *supra* note 81, at 107.

124. Joint Statement by President Clinton and Prime Minister Tony Blair of the United Kingdom (Mar. 14, 2000), https://www.govinfo.gov/content/pkg/WCPD-2000-03-20/pdf/WCPD-2000-03-20-Pg550.pdf.

125. HGP First Draft, *supra* note 114 (making all sequencing data available in the journal *Nature*); J. Craig Venter et al., *The Sequence of the Human Genome*, 291 SCI. 1304 (2001) (publishing only some sequencing data, with restrictions, as agreed to by the journal *Science*) [hereinafter Celera First Draft]. The HGP used DNA extracted from a set of volunteers to assemble their draft human genome sequence, so it represents an average, composite genetic profile. HGP Fact Sheet, *supra* note 112. The human genome sequence published by Celera is allegedly derived mostly from Venter's own DNA. Witkowski, *supra* note 120, at 786.

126. Waterston et al., *supra* note 39, at 3712; *see* Jeffrey A. Schloss, *How to Get Genomes at One Ten-Thousandth the Cost*, 26 NATURE BIOTECHNOLOGY 1113, 1113 (2008). Celera generated an ordered sequence of the human genome using whole-genome shotgun assembly in less than 1 year, with sequencing initiated on September 8, 1999 and assembly completed on June 25, 2000. Celera First Draft, *supra* note 125, at 1306. Notably, the whole-genome shotgun sequencing approach used by Celera relied upon preliminary data generated by the HGP effort, with hierarchical sequencing of bacterial artificial chromosome clones. Barranco, *supra* note 122.

127. Hood & Rowen, *supra* note 105, at 2.

128. *See In the Crossfire: Collins on Genomes, Patents, and 'Rivalry'*, 287 SCI. 2396 (2000) (transcribing an interview with Francis Collins, the head of the National Human Genome Research Institute and leader of the HGP).

In 2001, when the Project was nearing its completion, the National Human Genome Research Institute (NHGRI) began planning the next phase of genomics research: reducing the cost of sequencing the human genome to $1000.[129] Advances in Sanger sequencing technology during the HGP had already increased throughput and reduced costs from ten dollars to ten cents per base pair.[130] Five large centers emerged as leaders in genome sequencing throughout the HGP (the Wellcome Trust Sanger Institute, the Broad Institute, the Genome Institute of Washington University in St. Louis, the Joint Genome Institute, and the Whole Genome Laboratory at Baylor College of Medicine), which, together, paved the way for continued improvements in sequencing technologies.[131] Hoping to even further reduce the cost and time of sequencing, the NHGRI invested over $100 million in research grants dedicated to the development of new sequencing technologies.[132] Applicants developing sequencing-by-synthesis, sequencing-by-ligation, and nanopore approaches received the most grants, as NHGRI predicted these technologies to be most likely to "achiev[e] orders-of-magnitude improvements in sequencing."[133] And in a matter of years, sequencing-by-synthesis (used by Illumina) became the dominating approach in the NGS market.

C. PHASE 3: PRELIMINARY RESEARCH DRIVING KEY PRE-ILLUMINA ADVANCES IN NGS

As discussed in Section III.B *supra*, the HGP catalyzed several efforts to develop NGS. By the midpoint of the HGP, scientists in the United States and Europe had already established an array of startup companies that each sought to develop and sell the first NGS machine. Although Illumina—then called Solexa—later emerged as the winner of this race, technology from several different researchers, academic labs, and startup companies set the stage for the modern Illumina platform (**Table 2**). This Section outlines the early advances that inspired the three "massively parallel" elements of the Illumina platform.[134]

129. Schloss, *supra* note 126, at 1113.

130. *Id.*

131. Hood & Rowen, *supra* note 105, at 2.

132. Schloss, *supra* note 126, at 1114.

133. *Id.*; *Genome Technology Program*, NAT'L HUMAN GENOME RSCH. INST., https://www.genome.gov/Funded-Programs-Projects/Genome-Technology-Program#6 (last visited May 2, 2023).

134. *See* discussion *supra* Section II.B.1.

Table 2: List of critical people and institutions involved in the Illumina discovery story.

Phase	People	Institutions	Main Contributions	Relevant Active Years
Pre-Illumina	Frederick Sanger	Laboratory of Molecular Biology, United Kingdom	First-generation Sanger sequencing	1977
	Allan Maxam & Walter Gilbert	Harvard, United States	First-generation Maxam-Gilbert sequencing	1977
	Leroy Hood	Caltech, United States	Automation of Sanger sequencing	1980–1985
	Craig Venter	National Institutes of Health, United States	Celera Genomics, parallel HGP effort	1998–2003
	George Church	Harvard & Massachusetts Institute of Technology (MIT), United States	Multiplexing and solid support NGS	1984–1988
	Pascal Mayer	Serono Pharmaceutical Research Institute & Manteia Predictive Medicine, Switzerland	Bridge PCR clustering	1996–2004
	Pål Nyrén	Royal Institute of Technology, Sweden	Pyrosequencing	1986–1996
	Bruno Canard & Robert Sarfati	Pasteur Institute, France	Reversible terminator chemistry	1993–1994

Illumina	Shankar Balasubramanian & David Klenerman	University of Cambridge & Solexa, United Kingdom	Solexa, original platform	1997–2008
	John Milton	Solexa, United Kingdom	Medicinal chemistry at Solexa	2001–2008
	Nick McCooke & John West	Solexa, United Kingdom	Business development and expansion of Solexa	2001–2008
	Clive Brown, Klaus Maisinger & Tony Cox	Solexa, United Kingdom	Bioinformatics development at Solexa	2001–2008
	Sydney Brenner	Lynx Therapeutics, United States	Merged with Solexa	2005
	David Walt	Illumina, United States	Acquired Solexa	2007

1. *Solid Support Array*

As discussed in Section II.B.1.a) *supra*, the Illumina platform uses a solid support array to immobilize the DNA strands undergoing sequencing. Immobilization has been a component of the Illumina platform since it was first visualized in 1998.[135] Many different DNA sequencing researchers independently conceived of this same immobilization idea around the same time as Illumina researchers. In fact, so many groups simultaneously pursued the use of a solid support that it is difficult to identify who invented it. Indeed, the positional separation of DNA achieved by a solid support is the main feature that sets apart many NGS technologies from their "first-generation" analogs.[136]

George Church, a professor at Harvard University and MIT, was a key advocate for the use of solid support arrays to improve DNA sequencing—earning him a reputation as the "founding father of genomics."[137] Church was

135. *See* discussion *infra* Section II.B.1.

136. Sanger sequencing, and its many permutations, all separate DNA by size, only at the end of the sequencing process. *See* discussion *supra* Section II.A.3.

137. Stephanie Huie, *Dr. George Church, Founding Father of Genomics*, WINDWARD INST. (Apr. 15, 2020), https://www.thewindwardschool.org/the-windward-institute/the-beacon/article/~board/beacon-archives/post/dr-george-church-founding-father-of-genomics.

likely the first scientist to envision solid support immobilization in the context of DNA sequencing, as described in a 1984 publication[138] with Gilbert, one of the scientists who invented Maxam-Gilbert sequencing.[139] In that publication, Church and Gilbert disclosed the idea of DNA immobilization on nylon membranes.[140] Then, in a 1988 publication, Church and Stephen Kieffer-Higgins described the advantage of "multiplex" sequencing—mixing a series of different DNA samples together, and then sequencing the entire pool—as "greatest when the mixing occurs as early as possible and separation occurs as late as possible."[141] Church and Kieffer-Higgins envisioned a method of DNA sequencing with an initial multiplexing step and subsequent DNA separation based on nylon membrane immobilization.[142] Church continued to pursue NGS technologies, and eventually moved into nanopore-based sequencing—which launched a third generation of sequencing approaches.[143]

The original notion of immobilizing DNA by anchoring it against some other physical surface—not for the purposes of sequencing—traces back to Stephen Fodor's proposal for the "DNA chip" in the 1980s.[144] Inspired by computer chips, the DNA chip idea suggested the possibility of fixing DNA molecules to a small chip for analysis.[145] This concept may have been disclosed for the first time in 1994 in both a scientific publication[146] and a patent

138. George M. Church & Walter Gilbert, *Genomic Sequencing*, 81 PROCS. NAT'L ACAD. SCIS. 1991 (1984).

139. *See supra* text accompanying note 89.

140. *Id.* at 1992.

141. George M. Church & Stephen Kieffer-Higgins, *Multiplex DNA Sequencing*, 240 SCI. 185, 185 (1988).

142. *Id.*

143. *See* discussion *infra* Part V. In 1995, Church filed a patent application claiming a method of sequencing DNA anchored to an interface between two pools of media, U.S. Patent No. 5,795,782 (filed Mar. 17, 1995, issued Aug. 18, 1998). This patent is the first disclosure of George Church's nanopore sequencing idea, and was eventually licensed to Oxford Nanopore Technologies.

144. *See Stephen P.A. Fodor, Michael C. Pirrung, J. Leighton Read, Lubert Stryer (Affymax Research Institute, Palo Alto, USA)*, EUR. PAT. OFF., https://www.epo.org/en/news-events/european-inventor-award/meet-the-finalists/stephen-pa-fodor-michael-c-pirrung-j (last visited Mar. 13, 2024). For additional foundational research in this area, see, e.g., Robert L. Letsinger & V. Mahadevan, *Oligonucleotide Synthesis on a Polymer Support*, 87 J. AM. CHEM. SOC'Y 3526 (1965) (synthesizing DNA on a surface in the 1960s); M. D. Matteucci & Marvin H. Caruthers, *Synthesis of Deoxyoligonucleotides on a Polymer Support*, 103 J. AM. CHEM. SOC'Y 3185 (1981) (same but in the 1980s); *see also* Christine R. Laramy, Matthew N. O'Brien & Chad A. Mirkin, *Crystal Engineering with DNA*, 4 NATURE REVS. MATERIALS 201 (2019) (reviewing early work involving DNA attachment).

145. *Id.*

146. Jeffrey W. Jacobs & Stephen P.A. Fodor, *Combinatorial Chemistry—Applications of Light-Directed Chemical Synthesis*, 12 TRENDS BIOTECHNOLOGY 19 (1994).

application.[147] The possibility of then actually *sequencing* that immobilized DNA, in a manner distinct from that used by Church, Gilbert, and Kieffer-Higgins, was discussed by Andrei Mirzabekov in 1994.[148] And even earlier, another group of scientists filed a 1992 patent application on a method of amplifying DNA immobilized to a surface.[149] A 1995 scientific publication also disclosed attaching DNA to a surface.[150]

Tracking these early developments, a 2008 review noted that "[a]ll of the recently released, or soon-to-be-released, non-Sanger [NGS] commercial sequencing platforms . . . fall under the rubric of a single paradigm . . . [where] DNA (as a single molecule or in multiple copies) [is] *physically immobilized* on [an] array."[151] A 2010 review similarly identified "[a] common theme among NGS technologies is that the template is attached or immobilized to a solid surface or support."[152] The spatial separation achieved by physical immobilization of DNA provides unparalleled organizational power.

2. Bridge PCR Clustering

Again, as detailed in Section II.B.1.b) *supra*, the Illumina platform uses bridge PCR clustering to amplify template DNA strands, a unique approach that maintains physical anchoring to a solid support despite ongoing rounds of amplification. The amplification process produces groups of DNA molecules that Illumina terms "clusters." Illumina incorporated the bridge PCR clustering technology into its platform in 2004.[153] However, Pascal Mayer proposed an early version of bridge PCR clustering in 1996.[154] Back then, researchers called the characteristic Illumina clusters "colonies."

Mayer's work was pivotal to the eventual success of the Illumina platform. From 1996 to 2000, Mayer worked as a scientist in Geneva at the Biomedical Research Institute of GlaxoWellcome, which became the Serono

147. European Patent No. 0,476,014 (file June 7, 1990) (granted Aug. 31, 1994).

148. Andrei D. Mirzabekov, *DNA Sequencing by Hybridization—a Megasequencing Method and A Diagnostic Tool?*, 12 TRENDS BIOTECHNOLOGY 27 (1994).

149. U.S. Patent No. 5,616,478 (filed Oct. 26, 1992) (granted Apr. 1, 1997).

150. Mark Schena et al., *Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray*, 270 SCI. 467 (1995).

151. Jay A. Shendure et al., *Overview of DNA Sequencing Strategies*, 81 CURRENT PROTOCOLS MOLECULAR BIOLOGY 1, 4 (2008) (emphasis added).

152. Metzker, *supra* note 42, at 32.

153. *See* discussion *infra* Section III.D.

154. Lucie Aubourg, *French Scientist Recognized for Rapid DNA Sequencing Technique Key in COVID Fight*, PHYSORG (Sept. 9, 2021), https://phys.org/news/2021-09-french-scientist-rapid-dna-sequencing.html.

Pharmaceutical Research Institute.[155] In these years, he developed the "colony" approach to DNA sequencing,[156] first disclosing the idea in two 1997 patent applications.[157] A slide from a 1998 Mayer presentation shows the proposed DNA colonies in the distinctive "bridged" position—now synonymous with Illumina sequencing.[158] At the turn of the century, Mayer pushed for Serono to launch an independent startup company called Manteia Predictive Medicine to continue development of this new, "massively parallel" DNA sequencing approach.[159] Manteia further optimized the colony technology, as publicly disclosed in a 2003 presentation.[160] But the company remained unable to actually generate complete sequencing data and, that year, took half of their staff members off the project.[161] Manteia's investors lost interest and eventually withdrew funding.[162] The reduced investor interest motivated Manteia to eventually license the colony patents to the company established by Balasubramanian and Klenerman.[163]

3. SBS Read Generation

Finally, as described in Section II.B.1.c) *supra*, the Illumina platform uses a technology for read generation called SBS.[164] SBS is "massively parallel" because it allows for bases to be called at the moment of nucleotide incorporation, rather than indirectly at the end of the sequencing reaction. For Illumina, SBS is facilitated by the reversible terminator chemistry on the nucleotides that are incorporated, which was likely first outlined by the Solexa researchers in a 1998 patent application.[165] However, other scientists

155. Pascal Meyer, LINKEDIN, https://fr.linkedin.com/in/pascal-mayer-6b652a13 (last visited Oct. 23, 2022).

156. Aubourg, *supra* note 154.

157. Pascal Mayer, *Breakthrough Prize 2022: Behind Every Success Story There Are Great Teams*, LINKEDIN (Sept. 9, 2021), https://www.linkedin.com/pulse/breakthrough-prize-2022-behind-every-success-story-great-pascal-mayer.

158. Pascal Mayer, *A Very Large Scale, High Throughput and Low Cost DNA Sequencing Method Based on a New 2-Dimensional DNA Auto-Patterning Process*, Presentation at the Fifth International Automation in Mapping and DNA Sequencing Conference, St. Louis, MI (Oct. 7-10, 1998) (depicting an early variant of "colony" DNA sequencing on slides 3-5).

159. Mayer, *supra* note 157.

160. Manteia Predictive Medicine, Corporate Presentation (Sept. 2003), https://www.slideshare.net/pascalmayer/manteia-non-confidentialpresentation200309 (depicting a bridge PCR clustering pipeline on slides 19-23).

161. Barry Whyte, *Once in a Generation: Pascal Mayer and the Birth of a Billion-Dollar Industry*, 1 GEN BIOTECHNOLOGY 49, 55 (2022).

162. *Id.* at 56.

163. Mayer, *supra* note 157; *see* discussion *infra* Section III.D.

164. *Explore Illumina sequencing technology*, *supra* note 69.

165. WO 2000/006770 (filed July 30, 1999, claiming priority date of July 30, 1998).

developed the idea of SBS well before Solexa developed their reversible terminator version.

A 1985 patent, filed by Robert Melamede, was likely the earliest discussion of SBS as a DNA sequencing strategy.[166] The specification highlighted the labor-intensive methodology of Maxam-Gilbert and Sanger sequencing, even with automation, and proposed DNA sequencing without radioactivity or gel electrophoresis.[167] Without much detail as to the method, Melamede described that detection of nucleotide incorporation might occur as part of the sequencing process, based on a decrease in nucleotide absorbance.[168]

In parallel, without knowledge of the Melamede '849 patent from 1985,[169] Pål Nyrén conceived of the SBS idea "[o]ne late afternoon in the beginning of January 1986, bicycling from the lab over the hill to the small village of Fullbourn."[170] Nyrén described SBS as "follow[ing] the activity of DNA polymerase during nucleotide incorporation into a DNA strand."[171] However, rather than Melamede's method of monitoring nucleotide absorbance, or the eventual Illumina method of monitoring fluorescence, Nyrén envisioned a method of monitoring pyrophosphate release. Pyrophosphate synthesis occurs naturally as DNA strands elongate,[172] so rather than adding fluorescent tags to nucleotides, Nyrén suggested a sequencing method that continuously monitored naturally-occurring pyrophosphate production throughout DNA replication.[173] However, given several technical issues and other complications, almost a decade passed before Nyrén published a workable design for this approach.[174] Nyrén's method—termed *pyrosequencing*—increased the sensitivity of Melamede's SBS proposal by 100–1000-fold, overcoming a critical

166. U.S. Patent No. 4,863,849 (filed July 18, 1985) [hereinafter the '849 patent].

167. *Id.*

168. *See id.*

169. *See* Pål Nyrén, *The History of Pyrosequencing*, *in* PYROSEQUENCING PROTOCOLS 1, 1–2 (Sharon Marsh ed., 2007) ("Much later, I learned that Bob Melamede, whom I met in Stockholm in 1997, had described the general principles of DNA sequencing-by-synthesis in a previously obtained patent.").

170. *Id.* at 2.

171. *Id.*

172. Pyrophosphate is a transient molecule that is released naturally at the incorporation of each nucleotide into growing DNA strands, based on ATP hydrolysis by DNA polymerase. Jithesh Kottur & Deepak T. Nair, *Pyrophosphate Hydrolysis is an Intrinsic and Critical Step of the DNA Synthesis Reaction*, 46 NUCLEIC ACIDS RSCH. 5875 (2018).

173. *See* Nyrén, *supra* note 169, at 1–2.

174. Mostafa Ronaghi et al., *A Sequencing Method Based on Real-Time Pyrophosphate*, 281 SCI. 363 (1998). The approach adds luciferase into the sequencing mixture, so that luminescence can be monitored as a readout for pyrophosphate production during nucleotide incorporation. *Id.* The different nucleotides are distinguished based on the intensity of the luminescence signal emitted. *Id.*

impediment of the approach as previously articulated.[175] In 2005, pyrosequencing became the basis of the first commercialized NGS platform: the 454 Life Sciences system.[176]

Researchers also considered other elements of reversible terminator chemistry prior to the Solexa idea. The first mention of reversible terminator chemistry as the backbone of SBS was possibly in an unpublished 1993 French patent application.[177] The inventors, Bruno Canard and Robert Sarfati, filed a few other patent applications in this family, but none issued.[178] Then, in a 1994 scientific publication, Canard and Sarfati described their synthesis of modified nucleotides for DNA sequencing.[179] Their idea involved protecting the 3'-end of the extending DNA molecule (where Illumina uses a 3'-*O*-azidomethyl blocking group) with a blocking group, chemically or enzymatically removing that blocking group, and regenerating a free 3'-hydroxyl group to resume strand elongation.[180] This method is quite similar to the SBS read generation method now used in the Illumina platform. However, the specific chemistry proposed by Canard and Sarfati was distinct from that of the Solexa team.[181] Canard and Sarfati envisioned a different 3'-moiety for each distinct nucleotide and also incorporated the fluorescent label at the 3'-substituted end itself.[182]

A group at Columbia published another early conceptualization of the reversible terminator chemistry idea for DNA sequencing in 2003.[183] Reversible terminator chemistry has flourished over the years, with scientists continuing to explore 3'-*O*-blocked and 3'-*O*-unblocked terminators.[184]

---

175. *See* Nyrén, *supra* note 169, at 3–4 ("When I later met Bob [Melamede], he was very happy to hear that his sequencing-by-synthesis concept worked and that I had circumvented the problem of DNA polymerase-activity monitoring.").

176. Marcel Margulies et al., *Genome Sequencing in Microfabricated High-Density Picolitre Reactors*, 437 NATURE 376 (2005) (describing the pyrosequencing protocol out of Jonathan Rothberg's group at 454 Life Sciences, with a purported 100-fold increase in throughput compared to Sanger sequencing).

177. French Patent No. 2,703,052 (filed Mar. 26, 1993), https://patents.google.com/patent/FR2703052B1/en.

178. *See* https://globaldossier.uspto.gov/#/result/publication/FR/2703052/1.

179. Bruno Canard & Robert S. Sarfati, *DNA Polymerase Fluorescent Substrates with Reversible 3'-Tags*, 148 GENE 1, 1 (1994).

180. *Id.*

181. *See id.* at 2–3.

182. *Id.*

183. Zengmin Li et al., *A Photocleavable Fluorescent Nucleotide for DNA Sequencing and Analysis*, 100 PROCS. NAT'L ACAD. SCIS. 414 (2003).

184. Fei Chen et al., *The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology*, 11 GENOMICS, PROTEOMICS & BIOINFORMATICS 34 (2013).

D. PHASE 4: DISCOVERY OF THE NGS SOLEXA IDEA

Section III.C, *supra*, explained that scientists outside of the Solexa team also worked on the use of solid support arrays, bridge PCR clustering, and SBS read generation. And much of this preliminary research occurred before, or at least in parallel to, the evolution of the Illumina platform. But two professors working at the Chemistry Department at the University of Cambridge established the Illumina NGS system, as it exists today: Shankar Balasubramanian and David Klenerman. Their approach, as initially proposed and refined over several years, combines the three "massively parallel" elements outlined earlier: the solid support array, bridge PCR clustering, and SBS read generation.

The Balasubramanian and Klenerman method was unique compared to the work discussed in Section III.C, *supra*, along two axes. First, the Solexa team developed new ideas and approaches within each of the three sequencing elements. Second, and more critically, the Solexa team revolutionarily chose to combine all three elements—something no other company pursued at the time. These two features set Solexa's platform apart from its competitors. This Section describes the evolution of the Illumina platform, from the initial Solexa idea in 1998 to their first scientific publication in 2008.

In the late 1990s, with the HGP underway, Balasubramanian and Klenerman began working together.[185] The two professors met in 1997, when Balasubramanian—a biochemist studying DNA polymerase—needed help with laser excitation for an experiment during manuscript revisions[186] and sought help from Klenerman—a physical chemist with expertise in laser spectroscopy.[187] This initial collaboration sparked discussions about visualizing DNA polymerase while adding nucleotides in real-time, which combined both of their experimental areas.[188] Specifically, they wanted to capture the exact

185. Shankar Balasubramanian & David Klenerman, *Journeys of Discovery: Rapid Genome Sequencing*, YOUTUBE (May 18, 2021), .

186. Kevin Davies, *The Solexa Story*, BIOIT WORLD (Sept. 30, 2010), https://www.bio-itworld.com/news/2010/09/30/the-solexa-story (last visited Oct. 2, 2022).

187. *Id.*; Phil Prime, *The Award-Winning Researcher Behind Next Generation Sequencing*, CANCER RSCH. UK (Oct. 15, 2021), https://news.cancerresearchuk.org/2021/10/15/the-award-winning-researcher-behind-next-generation-sequencing/ (transcribing an interview with Balasubramanian).

188. Prime, *supra* note 187. This is essentially a version of SBS. But it is unclear whether Balasubramanian and Klenerman were aware of any other groups pursuing SBS research at this time. The Nyrén paper was released only in 1998, and before that, the only other report of SBS seemed to be in the 1985 Melamede '849 patent—and it is not unexpected that the Cambridge team would not have come across a United States patent when planning basic research projects. Some reports of the Solexa story seem to suggest that Balasubramanian and Klenerman independently conceived of their own SBS idea. Louise Walsh, *Journeys of Discovery*,

moment of nucleotide incorporation as it happened—not for the purposes of DNA sequencing, but simply to interrogate the enzyme kinetics of DNA polymerase. To this end, Balasubramanian and Klenerman envisioned the use of a solid support array to anchor a strand of DNA during nucleotide addition, and the use of fluorescent tagging to monitor the addition of nucleotides.[189]

In August 1997, the two decided to use their idea to for "massively parallel," NGS-type DNA sequencing. Balasubramanian, Klenerman, and their two postdocs (Mark Osborne, Colin Barnes) met at the Panton Arms, a pub near Cambridge where they routinely met to brainstorm ideas.[190] There, the group "saw the pieces of the jigsaw come together."[191] Their attempts to watch DNA be synthesized one molecule at a time, on a surface, had been repeatedly failing—they were constantly missing the actual moment of nucleotide incorporation.[192] To overcome this issue, the group proposed the idea of "watch[ing] lots of molecules in parallel at the same time."[193] This parallelization option had two implications: (1) probabilistically, they had be a better chance of "catching" the actual event of incorporation for at least one molecule; and (2) they could determine the sequence of all the DNA molecules on the surface, in parallel.[194] Sketching out the implications of this "massively parallel" sequencing plan on a piece of paper, the Cambridge team calculated that they could improve existing DNA sequencing technologies by up to 100,000-fold.[195]

The exact contours of the Panton Arms idea remain unknown, least of all the underlying biochemistry.[196] But a few months later, in November 1997, Balasubramanian and Klenerman approached venture capitalists at the Abingworth investment firm and presented their 100,000-fold "massively parallel" improvement plan.[197] After nine months of due diligence, in 1998, Abingworth provided seed funding for the team to form a company called

---

UNIV. CAMBRIDGE, https://www.cam.ac.uk/stories/journeysofdiscovery-rapidgenomesequencing (last visited Oct. 2, 2022) ("They realised [sic] that if they could *watch* the enzyme copying a genome then they were inadvertently also *reading* the genome. They had discovered a radically new way to sequence DNA").

189. Prime, *supra* note 187. It is unclear what their inspiration for this idea was—whether independent or based on the previous publications suggesting solid support arrays for sequencing.

190. Davies, *supra* note 186; Prime, *supra* note 187.

191. Balasubramanian & Klenerman, *supra* note 185.

192. *Id.*

193. Prime, *supra* note 187.

194. *Id.*

195. *Id.*; Balasubramanian & Klenerman, *supra* note 185.

196. *See* Prime, *supra* note 187.

197. Davies, *supra* note 186.

Solexa.[198] At this time, Osborne and Barnes were Solexa's only bench chemists.[199]

Solexa first publicly disclosed their sequencing plan in a PCT application filed in 1999 that claimed priority to an unpublished EPO application from 1998.[200] This publication detailed the combination of two out of three modern "massively parallel" elements of modern Illumina sequencing: a solid support array to which fragmented template DNA strands bind; and a roughly outlined sketch of fluorescent nucleotides for SBS (which would later become reversible terminator chemistry).[201] Critically, the publication lacked the bridge PCR clustering step, as the initial Solexa plan focused on *single-molecule* sequencing.[202] As described previously, DNA sequencing typically begins with some form of PCR amplification step to generate sufficient template for eventual base calling. Single-molecule sequencing omits the amplification step so that each individual fragment remains at low copy number.[203]

In the years after this early articulation, Solexa researchers focused primarily on refining the chemistry of their approach to SBS to further improve throughput and reduce technical complexity.[204] During these years, the company began to slowly grow. Additional funding from Abingworth allowed Solexa's facilities to move from the Chemistry Department at Cambridge to a lab at Chesterford Research Park.[205] Solexa hired four new employees: a research director (Harold Swerdlow), a CEO (Nick McCooke), a chief science officer (Tony Smith), a medicinal chemist (John Milton), and a bioinformatician (Clive Brown).[206]

---

198. *Id.*; Balasubramanian & Klenerman, *supra* note 185.

199. Davies, *supra* note 186; Balasubramanian & Klenerman, *supra* note 185.

200. Patent Appl. No. WO 2000/006770 (filed July 30, 1999, claiming priority date of July 30, 1998). The 1998 EPO patent application seems to be the first patent filed by Solexa.

201. *Id.* ("hybridising a polynucleotide molecule to its immobilised complement on the array . . . wherein each nucleotide triphosphate is conjugated at its 3' position to a different label capable of being characterised [sic] optically, determining which label (and thus which nucleotide) has undergone the polymerisation [sic] reaction, and removing the label").

202. *See What happened to Illumina's Single Molecule Sequencing Or Do You Remember Solexa's SMA-seq?*, ENSEQLOPEDIA, http://enseqlopedia.com/2012/08/what-happened-to-illuminas-single-molecule-sequencing-or-do-you-remember-solexas-sma-seq/ (last visited Oct. 2, 2022); Simon Bennett, *Solexa Ltd*, 5 PHARMACOGENOMICS 433, 434–35 (2004).

203. John F. Thompson & Patrice M. Milos, *The Properties and Applications of Single-Molecule DNA Sequencing*, 12 GENOME BIOLOGY 1 (2011).

204. *See* Davies, *supra* note 186.

205. *Id.*; *Solexa: Second-Gen Genetic Sequencing*, UNIV. CAMBRIDGE (July 13, 2015), https://www.enterprise.cam.ac.uk/case-studies/solexa-second-generation-genetic-sequencing/; *History of Sequencing by Synthesis*, ILLUMINA, https://www.illumina.com/science/technology/next-generation-sequencing//llumine-sequencing-history.html (last visited Oct. 2, 2022).

206. Davies, *supra* note 186.

Milton focused on redesigning the chemistry of the sequencing platform, modifying aspects of both the solid support array and the SBS reversible terminator nucleotides.[207] In parallel, Solexa diversified their patent portfolio by filing patents on all different aspects of the sequencing technology.[208] However, the actual detection of the fluorescent tags on the SBS nucleotides was failing.[209] Constrained by the single-molecule sequencing format, each incorporated fluorescent nucleotide on a single template strand could not yield a light signal intense enough for accurate base calling.[210] And the template strands themselves, though successfully adhered to the solid support array on one end, kept falling over, rather than standing straight up, which sterically prevented the growth of a complementary strand.[211]

Luckily, Manteia already had a solution to this steric hindrance problem. The Manteia team had visited the Solexa facilities in 2003 and noted Solexa's strength in reversible terminator chemistry but absence of actual sequencing data.[212] Given that Manteia was in a similar position, but with a different strength—functional cluster technology—Mayer found Solexa attractive.[213] So, in 2004, the Manteia group agreed to sell their clustering technology patents to Solexa.[214] Many view the Manteia acquisition as rescuing the Solexa platform: the low signal intensity from *un-amplified* DNA molecules was an insurmountable hurdle of the single-molecule format.[215] With the new possibility of clustered, bridge PCR amplification, Solexa almost instantly sequenced their first genome in 2005.[216] Their target was the bacteriophage phi X174 genome, which Sanger previously sequenced using the pre-Sanger sequencing "plus and minus" method.[217] Over a February weekend, Brown, along with two newly hired bioinformaticians (Klaus Maisinger and Tony

207. *Id.*

208. *See* discussion *infra* Section IV.B.2.

209. Davies, *supra* note 186.

210. *Id.*

211. *Id.* ("Klenerman tried putting a loudspeaker under the chip blasting high-frequency sound waves to make the DNA stand on end . . . that didn't work.").

212. *See* Whyte, *supra* note 161, at 56.

213. *Id.* at 57.

214. Davies, *supra* note 186.

215. *See* Clara Rodríguez Fernández, *The Man Behind Next-Generation Sequencing* (Mar. 11, 2019), https://www.labiotech.eu/interview/next-generation-sequencing-nick-mccooke/ (quoting McCooke, the former CEO of Solexa, stating that "the bridge amplification technology . . . from a Swiss company called Manteia . . . really saved the day . . . [i]f we hadn't been able to acquire that technology, the project would have failed").

216. *Id.*

217. Frederick Sanger et al., *Nucleotide Sequence of Bacteriophage phi X174 DNA*, 265 NATURE 687 (1977).

Cox), assembled the genome by short-read alignment, revealing more than 99.9% accuracy.[218]

To summarize, the Illumina sequencing platform combined three independent elements into a massively parallel process: (1) the use of a solid support; (2) the bridge PCR amplification of DNA fragments to generate clusters; and (3) the technique of SBS. All three elements represent remarkable innovations on their own, with each addressing important bottlenecks at stages of the sequencing process. However, as noted in Section III.C, *supra*, these inventions each existed in some primitive form before Solexa. Therefore, Solexa's key "inventive steps" were: (1) refining the biochemical techniques introduced into each of the three elements, especially in the reversible terminator chemistry for SBS; and (2) uniquely choosing to combine all three features, especially in licensing the bridge PCR clustering technology from Manteia.

E. PHASE 5: EXPANSION AND COMMERCIALIZATION OF SOLEXA/ILLUMINA

The Solexa team had the technological capacity to dominate the NGS market as early as 2005. But to take the next steps in expanding the company and commercializing their NGS method as a platform technology, Solexa required more robust business development practices. Key to this phase of the Illumina story is John West, the CEO of Solexa from 2004 to 2007.[219] West previously worked at Applied Biosystems on automated Sanger sequencing technology, and aspired to turn Solexa into an international company for its next phase of development.[220] Soon after becoming CEO, West negotiated a merger with Lynx Therapeutics, a biotechnology company based in California.[221] Lynx was led by Sydney Brenner, who for a time worked at the Laboratory of Molecular Biology (affiliated with the Medical Research Council), and then moved to California to establish the Molecular Sciences Institute. Lynx was also working on NGS technologies, but focused on a bead-based immobilization approach (rather than a solid support array).[222] Brenner

218. Davies, *supra* note 186.

219. John West, *"A Celebration of Solexa" – A Short Tour of DNA Sequencing History*, https://www.personalis.com/a-celebration-of-solexa-a-short-tour-of-dna-sequencing-history/ (last visited Oct. 2, 2022).

220. *Id.*; Davies, *supra* note 186.

221. Davies, *supra* note 186.

222. Nicholas Wade, *SCIENTIST AT WORK: SYDNEY BRENNER; A Founder of Modern Biology Shapes the Genome Era, Too*, N.Y. TIMES (Mar. 7, 2000), https://www.nytimes.com/2000/03/07/science/scientist-work-sydney-brenner-founder-modern-biology-shapes-genome-era-too.html.

has stated that he began thinking about the Lynx platform in the late 1980s while still in the United Kingdom, but failed to raise the requisite investment to begin the project.[223]

With the Solexa-Lynx merger in March 2005, just one month after Solexa had sequenced its first genome, Solexa became an international public company on NASDAQ.[224] And in 2006, Solexa launched its first sequencer machine, called the 1G Genome Analyzer (GA).[225] This machine could sequence one gigabase of data per run, and an entire genome for $100,000 in three months.[226] The Solexa GA was almost the first high-throughput sequencer on the market, but missed the mark by just one year. The 454 Life Sciences GS20 machine, which used automated Sanger sequencing, was released in 2005.[227]

Across the pond, Illumina had been slowly growing since 1998, founded based on the microarray platform developed by David Walt at Tufts University.[228] The company was focusing on gene expression analysis, using bead-based technology.[229] However, in January 2007, Illumina entered the NGS market with the acquisition of Solexa for approximately $650 million.[230] Again, West was critical in negotiating this acquisition.[231] And in the first month after the acquisition, Illumina sold twelve Solexa GA machines; by the end of 2007, Illumina installed more than 200 GAs in various institutes.[232] Genome sequencing centers and core facilities became more popular, as NGS became more economically feasible. In 2008, Illumina introduced an updated, increased-throughput sequencing machine ("GAII").[233] Finally, in November 2008, Balasubramanian, Klenerman, and approximately 100 other authors

223. William A. Wells, *Life After Worms, Lynx Therapeutics, Inc.*, 7 CHEMISTRY & BIOLOGY R191, R191 (2000) ("Most people thought it was too risky . . . [t]hey were pretty much right.").

224. *History of sequencing by synthesis*, *supra* note 205.

225. *Id.*

226. *Id.*

227. *Saying goodbye to 454: how to choose your next NGS platform*, BITESIZEBIO (Dec. 17, 2014), https://bitesizebio.com/22147/saying-goodbye-to-454-how-to-choose-your-next-ngs-platform/ (last visited Oct. 24, 2022).

228. *See generally* Deirdre Bradford Parsons, *Seminal Genomic Technologies: Illumina, Inc. & High-Throughput SNP Genotyping Beadarray Technology* (Nov. 19, 2007) (M.S. thesis, Duke University) (outlining a history of the Illumina company, with a focus on their pioneering microarray system).

229. *See* Davies, *supra* note 186.

230. *Id.*

231. West, *supra* note 219.

232. *Id.*

233. *Id.*

published the use of Solexa (now Illumina) sequencing technology for the first time in *Nature*.[234]

Since the initial Panton Arms proposal by Balasubramanian and Klenerman, Illumina has added another 10,000-fold increase in throughput with further optimization.[235] Now, the technology can sequence a human genome for $1,000 in one day, and the life sciences applications of Illumina sequencing go far beyond the basic research purposes that the Cambridge group initially envisioned.[236]

IV. INNOVATION DRIVERS IN THE DEVELOPMENT OF ILLUMINA NGS

Part III, *supra*, described the development history of the Illumina NGS platform, distilled into five distinct phases (**Figure 3**). One practical, motivating goal persisted throughout this story: the desire to facilitate faster and cheaper sequencing of human genomes. However, many additional sources of motivation, features of scientific research, and intellectual property protection strategies catalyzed the development of the Illumina NGS platform. The five stages of development illustrate many different innovation drivers that propelled the Illumina story forward.

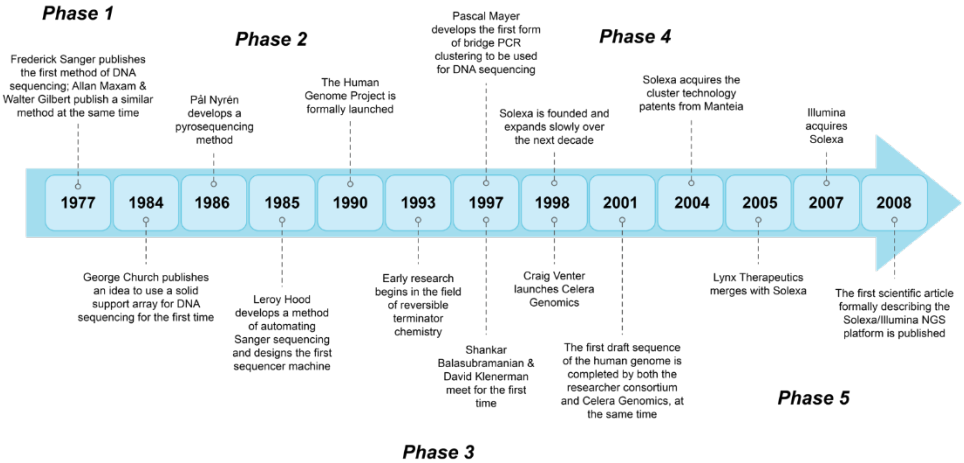
This Part describes and analyzes the relevant motivational factors for the scientists and institutions, tracking with the chronological development of Illumina's NGS platform technology. At the highest level, the innovation drivers fall into two distinct categories: an initial foundational period rooted in scientific fascination and altruism, and then a business development period characterized by intellectual property protection and commercialization.

234. Bentley et al., *supra* note 64, at 59.

235. Balasubramanian & Klenerman, *supra* note 185.

236. *Id.*; *see* Walsh, *supra* note 188 ("Their hopes for the technology have been exceeded over and over again.").

Figure 3: Timeline of NGS development.



A. FOUNDATIONAL INNOVATION DRIVERS

While several innovation drivers contributed to the remarkable innovation of the Illumina NGS platform, scientific curiosity, altruism, public funding sources, academic recognition, and serendipity motivated researchers to establish a foundation from which the Solexa technology could grow.

1. *Initial Scientific Curiosity and a Lack of Patent Protection*

In the early stages of the Illumina discovery story, scientists seemed largely motivated by scientific curiosity, rather than commercialization or intellectual property protection. As discussed in Section III.A, *supra*, the first two methods of DNA sequencing were published almost simultaneously in 1977: Maxam-Gilbert sequencing at Harvard University,[237] and Sanger sequencing at the Laboratory of Molecular Biology in Cambridge.[238] The scientists at this stage seem to have been motivated primarily by basic scientific curiosity. DNA sequencing began as a way of answering scientific questions to supplement basic molecular biology research.

Neither foundational sequencing method was patented, although the landscape at the time suggests that Maxam-Gilbert sequencing could have been.[239] At the time, U.S. universities and scientists rarely sought molecular biology patents based on academic research, though some existed: for

237. Maxam & Gilbert, *supra* note 89 and accompanying text.

238. Sanger et al., *supra* note 88.

239. *See* Holman, *supra* note 90, at 1054; Robert Cook-Deegan & Christopher Heaney, *Patents in Genomics and Human Genetics*, 11 ANN. REV. GENOMICS & HUM. GENETICS 383, 392 (2010).

example, Purdue University held a patent (granted in 1973) on a rudimentary method of DNA sequencing and Johns Hopkins University held a patent (granted in 1977) on pro-inflammatory nucleic acids.[240] But the decision for Maxam and Gilbert to forgo patent rights was not atypical; patent protection for U.S. university inventions only became normal practice after the 1980 Bayh-Dole Act and the Cohen-Boyer patent.[241] Indeed, Gilbert described having the sequencing invention in hand two years before publishing it in 1977, but noted that he had not considered patenting at all, given the ethos of the time.[242] Gilbert even prepared handouts on how to perform sequencing and distributed them to other scientists to encourage use—activity that would limit potential patent rights[243]—because he considered it merely a "basic research method[]."[244] Gilbert stated that his perspective on patenting was only shifted in the late 1970s, leading up to the Bayh-Dole Act, which he described as the government instructing university scientists to file patents for commercialization purposes.[245]

Sanger also forewent patent rights to his method of DNA sequencing. He received funding through the Medical Research Council, which at the time did not allow funded researchers to seek patent protection.[246] In 2001, during his Nobel Prize interview, Sanger expressed that even without this policy, he would not have wanted to patent the approach because he "wouldn't want to keep [his] work secret."[247] This anti-patent perspective, at the time, was not uncommon among basic science researchers.

---

240. *See* Cook-Deegan & Heaney, *supra* note 239, at 392; U.S. Patent No. 3,730,844 (filed Aug. 27, 1971); U.S. Patent No. 4,024,222 (filed Oct. 30, 1973).

241. *See* Cook-Deegan & Heaney, *supra* note 239, at 392; U.S. Patent No. 4,237,224 (filed Nov. 4, 1974).

242. *Walter Gilbert Interview*, NOBEL PRIZE (Mar. 22, 2009), https://www.nobelprize.org/prizes/chemistry/1980/gilbert/interview/.

243. *Id.*

244. Cook-Deegan & Heaney, *supra* note 239, at 392.

245. Gilbert expressed that even if he had patented his method of chemical DNA sequencing, it would not have made a difference, as the patent would have expired before the HGP (an entirely non-commercialized effort) was even completed. *See Walter Gilbert Interview*, *supra* note 242 (stating that he learned that patenting was necessary for commercialization, and that he "stopped thinking of patenting as an evil, [and] began to realize it as a social benefit," given the public disclosure aspect). In 1978—one year after the Maxam-Gilbert sequencing publication—Gilbert filed his first patent on bacterial-mediated production of insulin. U.S. Patent No. 4,411,994 (filed June 8, 1978). This patent was later licensed to Biogen. Marjorie Sun, *Biogen Pays High Price for Harvard Patent*, 222 SCI. 1309 (1983).

246. *Frederick Sanger Interview*, *supra* note 4.

247. *Id.* ("I wouldn't have wanted to [take patents] I don't think because I wouldn't want to keep my work secret . . . I don't think it would be fair.").

Overall, it appears that the foundational first-generation work needed to set the stage for later development in NGS did not specifically require patent protection or commercialization potential as incentivizing forces. The scientists appear to have been driven by a primary interest in molecular biology and biochemistry research.

2. *Altruism and Commercialization Antagonism in the Human Genome Project Era*

Two competing perspectives emerged in the second phase of the Illumina story, when scientists around the world recognized the potential impact of sequencing the first human genome.

Some scientists saw the HGP as a step towards revolutionizing healthcare and theorized that global access to a complete reference human genome sequence would be invaluable for modern medicine. Many felt that the HGP was the first step towards a world in which patients could routinely have their genomes sequenced in doctor's offices. In keeping with this altruistic ideology, researchers participating in the HGP received noncommercial, public funding through the NIH and the Department of Energy in the United States and through the Medical Research Council and the Wellcome Trust in the United Kingdom.[248] Ultimately, the goal of the HGP was to produce a human genome sequence that was entirely unpatented and freely accessible.[249]

Other scientists saw the human genome sequence as a potential trove of marketable data. The competing sequencing effort launched by Venter's company, Celera, received private funding and sought to generate a proprietary human genome sequence with patent protection on as many as 6,500 genes.[250] Although Venter's patent strategy fell through, Celera briefly licensed its genomic data to various institutions.[251]

Both the HGP and Celera completed their respective human genome sequences at the same time in 2001, with the HGP publishing a complete sequence in *Nature* and Celera publishing a partial one in *Science*. Researchers initially considered the Celera version "superior to the public [HGP] sequence"[252]—likely due to Celera's use of shotgun assembly[253]—allowing Celera to transiently profit from their data.[254] However, over the next year, the

248. HGP Fact Sheet, *supra* note 112.
249. *See id.*
250. Jeff Fox, *Sequencing, Patenting Surge*, 17 NATURE BIOTECHNOLOGY 1148, 1148 (1999).
251. Jim Kling, *Where the Future Went*, 6 EMBO REPS. 1012, 1012 (2005).
252. *Id.*
253. *See* discussion *supra* Section III.B.
254. Heidi L. Williams, *Intellectual Property Rights and Innovation: Evidence from the Human Genome*, 121 J. POLITICAL ECON. 1, 2 (2010).

costly Celera sequence lost its appeal, as the publicly accessible HGP data increased in quality.[255] With this, Celera bowed out of the genomics race to instead pursue drug discovery and diagnostic development.[256] By 2003, Celera turned their sequence data entirely over to the public domain.[257]

The HGP and Celera outlooks fundamentally split along ideological lines. Most of the scientific community was averse to the idea of commercializing DNA, viewing DNA as the most fundamental building block of life, generating research that "touches human lives directly," with "human beings or their cells . . . [as] the objects of [that] research."[258] For these scientists, gene patents were extremely controversial.[259] The Supreme Court initially enabled these patenting efforts in the 1980 *Diamond v. Chakrabarty* decision by allowing inventors to patent living organisms.[260] Based on this decision, Venter sought patent protection for expressed sequence tags (ESTs)—fragments of cDNA, not whole genes—in the early 1990s.[261] Many disapproved of this trajectory, given the concern that an EST patent landscape would restrict future research.[262] And when Venter pushed for a similar path during the HGP, much of the scientific community was even more troubled.[263] These years also saw the attempted patenting of the *BRCA1* and *BRCA2* genes, which was met with "overwhelmingly negative" public perception.[264] Some expressed concern that as many as 20% of human genes had been patented as of 2005,[265] possibly precluding the development of future sequencing technologies and diagnostic tests.[266]

---

255. *Id.* The HGP data was uploaded to the National Center for Biotechnology Information database.

256. *Id.*

257. *Id.*

258. Cook-Deegan & Heaney, *supra* note 239, at 388.

259. *Id.* at 389.

260. *See* Diamond v. Chakrabarty, 447 U.S. 303 (1980).

261. Daniel J. Kevles & Ari Berkowitz, *The Gene Patenting Controversy: A Convergence of Law, Economic Interests, and Ethics*, 67 BROOK. L. REV. 233, 235 (2001).

262. *Id.* at 237–39.

263. *Id.* at 245–48.

264. Cook-Deegan & Heaney, *supra* note 239, at 389–90.

265. *See* Kyle Jensen & Fiona Murray, *Intellectual Property Landscape of the Human Genome*, 310 SCI. 239 (2005).

266. Christopher M. Holman, *Debunking the Myth that Whole-Genome Sequencing Infringes Thousands of Gene Patents*, 30 NATURE BIOTECHNOLOGY 240 (2012). This turned out to not be entirely true. The substantive patentability doctrines, even for pre-*Myriad* gene patents, dictated sufficient specificity in claim language. With this, many of the alleged 20% of human gene patents were fairly narrow in scope, suggesting that whole-genome sequencing was never at risk of infringing on thousands of gene patents. Holman argues that these fears were based on a misinterpretation of the Jensen & Murray article. *Id.*

In the end, however, a perspective more aligned with the Bermuda Principles—which implicitly denounced gene patenting—won out. Post-*Chakrabarty* case law restricted patent eligibility for biological inventions,[267] the public increasingly viewed Venter as a villain,[268] and a global distaste for gene patents grew.[269] The idea that a company might profit from DNA sequence data eventually dissipated, leaving only DNA sequencing *instruments* up for patent protection.

Thus, overall, it seems that the central motivating factor for sequencing technology development during the HGP era was altruism. To a lesser extent, practical considerations were also likely important. In addition to the desire to secure a full-length, globally available human genome sequence, the HGP researchers also innovated out of frustration with the slow speed of the existing sequencing methods. Indeed, despite its eventual success, the HGP dramatically revealed the prohibitive costs associated with Sanger sequencing technologies, even once automated.

3. *Academic Recognition with Science and Technology Prizes*

In addition to altruism, professional recognition also likely motivated NGS researchers. Many knew that improving DNA sequencing technologies would provide great social value, but not everyone agreed with the level of importance. Anecdotally, in the early 2000s—when many different sequencing startups were in the early stages of development—high impact journals accepted many publications related to DNA sequencing development. This likely incentivized further research in this area.

Prize-awarding bodies have now recognized many of the discoveries associated with the development of the Illumina platform. For example, in 1980, Sanger, Gilbert, and Paul Berg jointly received the Nobel Prize in Chemistry for Sanger and Gilbert's DNA sequencing research and Berg's recombinant DNA studies.[270]

---

267. *See, e.g.*, Shahrokh Falati, *Patent Eligibility of Disease Diagnosis*, 21 N.C. J.L. & TECH. 63 (2020).

268. *See, e.g.*, Witkowski, *supra* note 120, at 786 (recounting the various epithets ascribed to Venter, including "maverick, publicity hound, risk-taker, brash, controversial, genius, manic, rebellious, visionary, audacious, arrogant, feisty, determined, provocative").

269. Cook-Deegan & Heaney, *supra* note 239, at 390–96.

270. *The Nobel Prize in Chemistry 1980*, NOBEL PRIZE, https://www.nobelprize.org/prizes/chemistry/1980/summary/ (last visited Nov. 25, 2022). Gilbert later went on to pursue other commercial endeavors. *See* Kanigel, *supra* note 106. Sanger remained with the Medical Research Council, and personally disavowed patent protection and other forms of commercialization for the rest of his life on the bench. *See Frederick Sanger Interview*, *supra* note 4.

Other NGS researchers received prizes that were focused on the practical implications of NGS. Over the years, the scientific community began to feel that Nobel Prizes were too restricted to "basic" science discoveries.[271] To fulfill the perceived gap in awards for more "applied" technological research, Finland established the Technology Academy in 2003,[272] which now awards the Millennium Technology Prize to innovations that "promote the well-being of humankind and society," and are specifically "appli[ed] with global commercial viability."[273] The Nobel Prize and the Millennium Technology Prize each award approximately $1 million. Balasubramanian and Klenerman each received the 2020 Millennium Technology Prize for NGS technology.[274] Notably, however, this prize did not exist during the critical years of technological advancement in the Illumina discovery story (i.e., from 1998 to the early 2000s)—and likely did not serve as a motivating factor for the scientists.

Yet another scientific prize was established in 2013, called the Breakthrough Prize in Life Sciences.[275] With funding from Mark Zuckerberg, Priscilla Chan, Sergey Brin, Yuri Milner, and Anne Wojcicki,[276] the Breakthrough Prizes each offer $3 million—the largest scientific award currently available.[277] Balasubramanian, Klenerman, and Mayer jointly received the 2022 Breakthrough Prize in Life Sciences.[278] This award was one of the first explicit recognitions of Mayer's critical involvement in the Illumina discovery story, as his work at Manteia went relatively unrecognized for several years.[279]

---

271. *See DNA sequencing pioneers win 1mn euro tech 'Nobel' prize*, PHYSORG (May 18, 2021), https://phys.org/news/2021-05-dna-sequencing-1mn-euro-tech.html.

272. *Technology Academy Finland*, MILLENNIUM TECH. PRIZE, https://millenniumprize.org/about-us/in-english/ (last visited Nov. 25, 2022).

273. *Story*, MILLENNIUM TECH. PRIZE, https://millenniumprize.org/prize/story/ (last visited Nov. 25, 2022).

274. *2020 Next Generation DNA Sequencing*, MILLENNIUM TECH. PRIZE, https://millenniumprize.org/winners/next-generation-dna-sequencing/ (last visited Nov. 25, 2022).

275. *Life Sciences Breakthrough Prize*, BREAKTHROUGH PRIZE, https://breakthroughprize.org/Prize/2 (last visited Nov. 25, 2022).

276. Rory Carroll, *Breakthrough Prize announced by Silicon Valley entrepreneurs*, GUARDIAN (Feb. 20, 2013, 12:00 AM), https://www.theguardian.com/science/2013/feb/20/breakthrough-prize-silicon-valley-entrepreneurs.

277. J.P., *Take that, Alfred*, ECONOMIST (Feb. 20, 2013), https://www.economist.com/babbage/2013/02/20/take-that-alfred.

278. *Winners of the 2022 Breakthrough Prizes in Life Sciences, Fundamental Physics and Mathematics Announced*, BREAKTHROUGH PRIZE, https://breakthroughprize.org/News/65 (last visited Nov. 25, 2022).

279. Mayer, *supra* note 157; Aubourg, *supra* note 154. Mayer noted that his inspiration for the bridge PCR clustering idea came from his post-doctoral research at the University of

Finally, a unique form of prize recognition occurred in 2017, when Queen Elizabeth II knighted Balasubramanian. The English monarchy has routinely granted damehood and knighthood to people who make significant scientific contributions, including Isaac Newton, Tim Berners-Lee, Jane Goodall, and Sarah Gilbert.[280] Balasubramanian's knighthood was attributed to his "services to science and medicine."[281] He has independently received several other, smaller scientific awards over the years.[282]

4. Federal Funding for Early Academic Research

Most of the people involved in the Illumina discovery story began their careers as scientists working in universities or other academic settings (**Table 2**). Each person or team followed a similar trajectory in the competitive NGS era of the early 2000s. As each group made progress in the development of a marketable NGS platform, the group would typically establish a startup company and run the company in parallel with their academic research. For example, Hood (Applied Biosystems), Venter (Celera Genomics), Brenner (Lynx Therapeutics), and Balasubramanian and Klenerman (Solexa) all followed this path.

With this canonical structure, national governments funded the bulk of research and development in the early phases of the Illumina development story. The researchers who developed first-generation sequencing technologies (Sanger, Maxam, Gilbert, Hood) all received federal research grants in the United States and/or the United Kingdom. Sanger received funding from the Medical Research Council in the United Kingdom, which is similar to the NIH in the United States. Maxam, Gilbert, and Hood received NIH grants. In a 2002 lecture, Hood explained that he viewed venturing into commercialization as requiring an appreciation of "long-term vision and

---

Strasbourg and the University of Ottawa. Mayer, *supra* note 157. His original terminology for the technique—"colony" sequencing—was later adopted by Church and other scientists, who developed a similar immobilization-based sequencing technique that they termed "polony" sequencing. Jay Shendure et al., *Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome*, 309 SCI. 1728 (2005) (describing the polony protocol out of George Church's group at Harvard). This concept would eventually become a key component of third-generation sequencing. *See* discussion *infra* Part V.

280. *Scientists Who Have Received a CBE, OBE, MBE, Knighthood or Damehood*, GAZETTE, https://www.thegazette.co.uk/all-notices/content/103527 (last visited Nov. 25, 2022).

281. *Professor Sir Shankar Balasubramanian FRS*, ROYAL SOC'Y, https://royalsociety.org/grants-schemes-awards/career-pathway-tracker/shankar-balasubramanian/ (last visited Nov. 25, 2022).

282. *Honour for Trinity Fellow Professor Shankar Balasubramanian*, TRINITY COLL. CAMBRIDGE, https://www.trin.cam.ac.uk/news/honour-for-trinity-fellow-professor-shankar-balasubramanian/ (last visited Nov. 25, 2022).

potential" and that "it is often true that radical new opportunities can progress more effectively as new startups rather than through the licensing to preexisting companies."[283] Hood's development of the automated sequencer machines relied on National Science Foundation funding, which generated "one of the most outstanding [programs] ever funded by the federal government."[284]

The geographic localization of NGS research suggests government funding was a key innovation driver. Despite *global* investment into the HGP—and general, widespread interest in sequencing technology development—the participants in the Illumina development story operated mainly in the United States and the United Kingdom. One explanation for this is the unique culture surrounding DNA and molecular biology-based research in both geographical areas. Decades of research investment turned both regions of the world into epicenters of discovery and expertise.

In later stages of pre-NGS research, scientists including Venter, Church, Mayer, Nyrén, Canard, Sarfati, Balasubrmanian, and Klenerman, also received federal funding through government grant programs in the United States, Canada, and Europe. Balasubramanian and Klenerman, for example, received funding from the Biotechnology and Biological Sciences Research Council.[285] Eventually, however, private funding sources took over.[286]

5. *Serendipity in the Initial Solexa Idea*

The Solexa team credits the August 1997 meeting at the Panton Arms as being the moment that the NGS idea was truly launched. This meeting highlights an instance of serendipity in the NGS invention story. Balasubramanian and Klenerman began working together for a project entirely unrelated to NGS, with Balasubramanian simply seeking out Klenerman's laser spectroscopy expertise to help with paper revisions. And the initial focus of Balasubramanian and Klenerman's research was on understanding the enzyme kinetics of DNA polymerase, not designing a commercialized NGS platform. Only at the Panton Arms did the team decide to implement a parallelization

283. Leroy Hood, President and Director, Institute for Systems Biology, Commemorative Lecture for the 2002 Kyoto Prize in Advanced Technologies, My Life and Adventures Integrating Biology and Technology (2002).

284. *Id.*

285. *Blog: Into the Unknown – Why Blue-Sky Research is Vital for Scientific Breakthroughs*, MILLENNIUM TECH. PRIZE (Jan. 31, 2022), https://millenniumprize.org/news-articles/blog/into-the-unknown-why-blue-sky-research-is-vital-for-scientific-breakthroughs/ [hereinafter Balasubramanian & Klenerman Millennium Prize Blog].

286. *See* discussion *infra* Section IV.B.1.

approach in their experiments, creating the possibility for DNA sequencing as an application.

Even after Balasubramanian and Klenerman turned their focus to NGS technologies, they described their work as basic, "blue skies" research.[287] Like most of the scientists that contributed to NGS technologies, Balasubramanian and Klenerman focused on solving general problems of molecular biology and biochemistry, such as the functionality of DNA. The prospect of translating this fundamental work into a commercialized sequencing technology was not necessarily an initial motivator for many involved, including the two Solexa founders. As Balasubramanian and Klenerman have explained, they "were just following [their] curiosity about the molecular machines that nature uses to copy . . . DNA."[288] And in fact, much of the Solexa team lost interest in the project when its goals became more "clinical" or "applied" in nature. Brown and several other scientists left Solexa specifically when the "scientific and commercial priorities chang[ed]," with Brown stating that he "wouldn't have left had we not done a human genome."[289]

B. BUSINESS-ORIENTED INNOVATION DRIVERS

Once the Solexa scientists launched their NGS platform, a different set of innovation drivers began to take over in relative importance. Increased research and development costs for the growing Solexa NGS platform necessitated private funding sources, a robust patent portfolio, well-timed licensing, dedication to commercialization, and aggressive litigation.

1. Transitioning into Private Funding Sources

By the time Balasubramanian and Klenerman launched their collaboration, the innovative landscape propelling NGS researchers forward had shifted. As discussed in Section III.C, *supra*, the turn of the century saw the independent discovery and development of many similar DNA sequencing techniques, which would later fit together into the now-canonical four-step NGS platform. These new, exciting ideas all shared one critical feature: their increased cost, relative to first-generation sequencing methods. To expedite the sequencing process, scientists in the United States and the United Kingdom focused on pursuing new types of terminator chemistry and solid support array configurations, requiring significant resources for long-term optimization.

Because of the significant increase in research and development costs at this later stage of the Illumina development story, private funding sources

287. Prime, *supra* note 187; Balasubramanian & Klenerman, *supra* note 185.
288. Balasubramanian & Klenerman Millennium Prize Blog, *supra* note 285.
289. Davies, *supra* note 186.

became much more important for many companies. Indeed, part of Solexa's (now Illumina's) success likely arose from the early influx of funding the Solexa team received from Abingworth, an investment firm interested in supporting DNA sequencing research, among other life sciences technologies. At this time, many scientists around the world had founded sequencing-based startup companies, and investor conferences to recruit funding from firms interested in DNA sequencing were becoming quite common. As discussed in Section III.E, *supra*, Solexa deliberately recruited several business-oriented team members to attend these conferences and present data from their nascent sequencing platform. Compared to other participants at these conferences, the Solexa team offered a more functional platform at earlier stages—likely justifying the financial support they received.[290] Indeed, more recently, Mayer has expressed his view that the difference between Solexa and Manteia was that "Solexa had the confidence of their investors, something [Manteia] didn't have . . . [Solexa's investors] believed in the project and wanted it to happen; ours saw out-of-core cash burn and wanted to cease diverting their attention."[291]

2. *Developing a Strong Patent Portfolio*

Unlike the scientists in the "first-generation" DNA sequencing era, Balasubramanian and Klenerman sought patent protection almost immediately after they envisioned the Solexa idea, to lay a foundation for future commercialization. Balasubramanian initially pitched their idea to researchers participating in the non-commercialized HGP.[292] But because they had not yet developed the technology, the HGP had little use for their idea and the Cambridge team "couldn't think of any other way[, besides starting a company,] of pulling together the kind of resource[s] that [they] need[ed]."[293] Indeed, the team approached Abingworth mere months after the Panton Arms meeting and founded Solexa the very next year, filing their first patent in July.[294]

Strong patent protection remained core to the Solexa (now Illumina) story throughout its development. And now, Illumina owns patents on virtually every eligible aspect of their technology.[295] For example, the physical sequencer machines that house the sequencing reactions are protected by

290. *See id.*

291. Whyte, *supra* note 161, at 57.

292. Prime, *supra* note 187.

293. *Id.*

294. Davies, *supra* note 186; Patent Appl. No. WO 2000/006770 (filed July 30, 1999, claiming priority date of July 30, 1998).

295. *See Illumina virtual patent marking*, ILLUMINA, https://www.illumina.com/company/legal/patents.html (last visited Oct. 22, 2022).

patents on methods of imaging the growing DNA strands, the imaging system itself, and structural aspects of the flow cell surface.[296] Similarly, the ancillary biochemical components of the sequencing reactions are protected by patents on various polymerases and buffers.[297] Many of these features are outside of the scope of this Article. However, they are also not the "core" elements that make the Illumina platform "massively parallel." That is, if not for the patents on the three key Illumina elements, a competing sequencing company could likely construct an Illumina-style sequencer machine that functioned with some equivalence.[298] It is the protection over the solid support array, bridge PCR clustering, and SBS chemistry (specifically, the modified reversible terminator nucleotides) that ties up the Illumina platform among other types of NGS. Each Section *infra* traces the development of a patent portfolio for these three core elements, beginning at the turn of the century and persisting with continuation patents still being filed on each element today.

a) Solid Support Array

The use of a solid support array was always part of the Solexa idea, originating from its ancestral basic science project. Solexa's first publicly available patent application (filed in 1999) claimed "[a] device comprising an array of molecules capable of interrogation and immobilised [*sic*] on a solid surface . . . wherein each molecule is immobilised [*sic*] at one or more points, by specific interaction with the surface."[299] After this first disclosure, they patented several other permutations of the solid support concept, spawning several patent families on different array architectures. For example, the "Arrayed Biomolecules" U.S. patent family begins in 2001 with an application claiming a slightly narrower articulation of the solid support array, including the use of oligonucleotides anchored to the array to bind to target DNA molecules.[300]

Over time, the solid support array patents became more complex, requiring integration with other aspects of the platform—likely to overcome the patent novelty and nonobviousness hurdles. Now, Solexa's patents primarily focus on many of the secondary biochemical factors that support immobilization, e.g., the ligase enzymes used for adapter attachment, the

296. *See id.*
297. *See id.*
298. *See The Next Few Years in DNA Sequencing*, 41J BLOG (May 23, 2021, 3:53 AM), https://41j.com/blog/2021/05/the-next-few-years-in-dna-sequencing/.
299. Patent Appl. No. WO 2000/006770 claim 1.
300. U.S. Patent No. 6,787,308 (filed Jan. 30, 2001) (granted Sept. 7, 2004).

methods of adapter attachment, the adapters themselves, and the siderophores used to help with preparation.[301]

b) Bridge PCR Clustering

Manteia filed the first PCT applications on Mayer's bridge PCR clustering technology in 1997.[302] These early applications were directed to a method of DNA amplification that includes a possible bridge configuration, with claims specifying the degree of immobilization of different ends of DNA strands.[303] The "Method of Nucleic Acid Amplification" U.S. patent family begins in 2001, with an application describing the use of a solid support featuring bridge PCR amplification in moderate detail.[304] A similar 2003 patent application (granted in 2011) has equally broad claims, but also includes drawings indicating a more detailed conception of the bridge PCR technique.[305] After Manteia licensed these patents to Solexa,[306] Solexa filed continuation applications regularly until each family's expiration (occurring from 2018 to 2020), gradually narrowing claim breadth regarding the specifics of the PCR process on the solid support.[307]

Notably, another U.S. patent family, filed earlier than the Manteia patent portfolio, also discloses bridge PCR clustering. Researchers at Mosaic Technologies and the Whitehead Institute for Biomedical Research filed the first application was filed in 1994.[308] The inventors, Christopher Adams and Stephen Krohn, filed a series of continuation applications up until 2000.[309] However, they lacked a complete understanding of the clustering process—while it was clear that some degree of localized amplification was occurring,

301. *See, e.g.*, U.S. Patent No. 8,877,939 (filed Feb. 23, 2011) (granted Nov. 4, 2014) (claiming ligation methods); U.S. Patent No. 10,525,437 (filed Jan. 8, 2018) (granted Jan. 7, 2020) (claiming adapter layout on solid support array and sequences of those adapters).

302. Patent Appl. No. WO 1998/044151 (filed Apr. 1, 1998) (published Oct. 8, 1998); Patent Appl. No. WO 1998/044152 (filed Apr. 1, 1998) (published Oct. 8, 1998). Mayer noted that the initial concept was "separated in two distinct inventions at the demand of [GlaxoWellcome's] patent department." Mayer, *supra* note 157.

303. *See, e.g.*, Patent Appl. No. WO 1998/044151 claim 27.

304. U.S. Patent Appl. No. 2004/0096853 (filed Dec. 7, 2001) (published May 20, 2004).

305. U.S. Patent No. 7,985,565 (filed June 2, 2003) (granted July 26, 2011) (illustrating the amplification strategy in FIG. 1B).

306. *See* discussion *supra* Section III.D.

307. *See, e.g.*, U.S. Patent No. 9,593,328 (filed Jan. 20, 2015) (granted Mar. 14, 2017); U.S. Patent No. 10,370,652 (filed Feb. 22, 2016) (granted Aug. 6, 2019).

308. U.S. Patent No. 5,641,658 (filed Aug. 3, 1994) (granted June 24, 1997).

309. *See, e.g.*, U.S. Patent No. 6,468,751 (filed June 9, 2000) (granted Oct. 22, 2002).

they lacked the imaging data to support a determination of clustering.[310] Manteia later acquired this patent family and licensed it to Solexa.[311]

c) SBS Read Generation

As discussed in Section II.B.1.c), *supra*, reversible terminator nucleotides facilitate the SBS element of the Illumina platform. One of the first Solexa disclosures, a 2000 PCT application, claims a nascent conceptualization of reversible termination.[312] Although other researchers had suggested versions of SBS before,[313] the 2000 PCT application is one of the first publications to articulate an SBS method with an exogenous, removable label.[314] The only prior discussion of reversible terminator chemistry in the context of DNA sequencing was most likely in the 1994 Canard and Sarfati publication.[315] Notably, the early Solexa application did not focus specifically on the nucleotides. It covered almost the entire Solexa platform as envisioned at the time, with claims directed to a sequencing device, the immobilization of DNA strands on a surface, and the method of sequencing itself.

Soon after, however, Solexa began to file patents on the more individualized components of its technology, with an emphasis on the reversible terminator nucleotides. The "Modified/Labelled Nucleotides" U.S. patent family begins in 2002, with a patent application (granted in 2006) containing broad genus claims directed to a nucleotide with: (1) a "protecting group" attached to the deoxyribose sugar at the 2' or 3' oxygen atom; and (2) a "detectable label" attached to the base via a "cleavable linker."[316] These two elements correspond to the modern-day 3'-*O*-blocking group and the fluorescent dye, respectively.[317] At this stage, this patent family provided no details as to the chemical structures of the protecting group, the detectable

310. Whyte, *supra* note 161, at 54.

311. *See id.*

312. Patent Appl. No. WO 2000/006770 (filed July 30, 1999, claiming priority date of July 30, 1998); *see supra* text accompanying note 165.

313. *See* discussion *supra* Section III.C.3.

314. Patent Appl. No. WO 2000/006770 (claiming a method where "each nucleotide triphosphate is conjugated . . . to a different label . . . determining which label . . . has undergone the polymerisation [sic] reaction, and removing the label"). There is, at least, a single 1994 patent (claiming a 1989 priority date) that claimed a similar modified nucleotide, but integrated into an entire sequencing method: "optically-labeled derivatives of four nucleotide 4'-triphosphates . . . where said optically-labeled derivatives comprise a *blocking group at the 3' portion thereof*, said blocking group comprising *an optical label capable of being removed to expose the 3' portion thereof*" (emphasis added). U.S. Patent No. 5,302,509 (granted Apr. 12, 1994).

315. *See* Canard & Sarfati, *supra* note 179.

316. U.S. Patent No. 7,057,026 (granted June 6, 2006).

317. *See* discussion *supra* Section II.B.1.

label, or the cleavable linker, aside from a single dependent claim to the detectable label as a fluorophore.

In subsequent years, Solexa filed increasingly narrower continuation applications, swirling around the single reversible terminator species that Illumina uses today: a nucleotide with an azidomethyl 3'-*O*-blocking group.[318] In 2003, a PCT application claimed a nucleotide with a "removable 3'-OH blocking group" where the 3' carbon atom has an oxygen atom bonded to any one of several chemical moieties.[319] This application also specified some parameters for the detectable label and cleavable linker, specifically, that the label may be a fluorophore and the linker may be "acid labile, photolabile, or contain[] a disulfide linkage."[320] Another 2003 application (granted in 2008) claimed specific chemical structures for the cleavable linker.[321] Solexa continued this rigorous patent acquisition strategy for several years.[322] The most recent patent in this family, filed in 2020, claims a much narrower genus: a nucleotide with a 3'-*O*-azidomethyl group, a single cleavable linker structure family (benzene-based, but with several permutations possible), and a fluorophore tag.[323] Milton, the first medicinal chemist hired when Solexa began to expand, is an inventor on even the newest applications.

As suggested in Section III.C.3, *supra*, the Solexa researchers were not the first to propose SBS. They were also not the first to disclose the idea of reversible terminator chemistry in the context of nucleotides. For example, Andrew Hiatt and Floyd Rose filed a 1995 patent application that claimed nucleotides with 3' removable blocking moieties.[324] These two inventors hold several other early U.S. patents directed to critical chemical structures in the

318. Bentley et al., *supra* note 64, at 53. Notably, 3-*O*-azidomethyl nucleosides were reported in 1991, but with no detectable label added. Sergey Zavgorodny et al., *1-alkylthioalkylation of Nucleoside Hydroxyl Functions and Its Synthetic Applications: A New Versatile Method in Nucleoside Chemistry*, 32 TETRAHEDRON LETTERS 7593 (1991).

319. Patent Appl. No. WO 2004/018497 claims 1-5 (where claim 4 is specifically directed to an azidomethyl group) (filed Aug. 22, 2003).

320. *Id.*

321. U.S. Patent No. 7,414,116 (filed Aug. 22, 2003) (granted Aug. 19, 2008).

322. *See Illumina virtual patent marking*, *supra* note 295.

323. U.S. Patent No. 11,028,115 (filed Sept. 28, 2020) (granted June 8, 2021).

324. U.S. Patent No. 5,872,244 (filed June 7, 1995) (granted Feb. 16, 1999).

reversible terminator nucleotide space.[325] Solexa licensed this patent family in 2005.[326]

Overall, it seems that the Solexa researchers recognized, very early on, that they would incur substantial research and development costs in putting together a robust NGS platform. They then chose to invest considerable effort in establishing a robust patent portfolio, as one way of scaffolding around this goal. The contrast between this approach and that of the researchers dedicated to first-generation sequencing technologies is striking. Prior to the passage of the Bayh-Dole Act (and the Solexa collaboration), molecular biologists tended to eschew the notion of intellectual property protection. Indeed, Sanger sequencing existed for quite some time before researchers sought any patents. But Sanger sequencing was initially a low-cost method performed at a very small scale. As soon as researchers saw the potential in scaling up Sanger sequencing, they also recognized that it could become cost-prohibitive. Thus, Hood's pursuit of an automated sequencer machine—to speed up the sequencing process—almost immediately dovetailed with iterative patent filing. The Solexa researchers likely attracted more investor support than their competitors, and certainly than their first-generation sequencing predecessors, due to their robust intellectual property rights.

3. *Strategic Licensing*

During the business expansion phase of the Solexa/Illumina story, the company licensed the cluster technology patents from Manteia. This licensing deal likely rescued the Solexa platform, which was at a standstill and had failed to generate any meaningful sequencing data. Compared to other sequencing companies working on similar NGS platforms at the same time, this was a remarkably unique decision. When Solexa acquired the cluster patents from Manteia, the scientific team effectively abandoned the alternative flow cell configurations they were pursuing. Relative to other groups which were seemingly attached to specific pipeline components, the Solexa team prioritized only the *ultimate* output of an effective sequencing method—rather than certain features (e.g., their former flow cell configurations). Solexa's pivot to cluster technology reflects a prioritization of business goals, rather than scientific ideals. This approach can likely be attributed to the business leaders that Solexa hired in the early 2000s.

---

325. *See, e.g.*, U.S. Patent No. 5,763,594 (filed June 7, 1995) (granted June 9, 1998); U.S. Patent No. 5,808,045 (filed June 7, 1995) (granted Sept. 15, 1998); U.S. Patent No. 6,214,987 (filed June 7, 1995) (granted Apr. 10, 2001).

326. *Solexa Strengthens Patent Position in Next-Generation Genetic Analysis*, TECH. NETWORKS GENOMICS RSCH. (Nov. 9, 2005), https://www.technologynetworks.com/genomics/news/solexa-strengthens-patent-position-in-nextgeneration-genetic-analysis-209745.

4. *Commercialization Potential*

The prospect of being the first company to commercialize an NGS machine motivated Solexa and other startup companies during the early 2000s. Those involved in the Illumina NGS journey—from the initial Solexa idea at the Panton Arms to the currently available sequencing machines—repeatedly attribute the success of the platform to their timing, rather than a specific innovation. Klenerman has stated that he believes the Solexa platform is now used so widely simply because their team was one of the first to commercialize the technology.[327] Milton similarly expressed that "Illumina dominates [because] they got there first," highlighting that once a genome sequencing center implements a specific NGS company's machine, the center's staff adapt to that machine's methodology, and are less likely to switch to a different type of technology.[328] And Smith suggested that if the GA machines reached the market just two years later, the difference in competition would have been enormous.[329] Indeed, in the years that followed the entry of the Solexa sequencer machine onto the market, many competitor sequencing companies successfully assembled their own sequencers and launched them.[330] But none of those companies experienced the same astounding success as Illumina, and in the past decade, almost all of those later-launched machines have since been taken off the market.[331] Now, new competitors looking to enter the NGS market "take[] care to ease adoption by developing conversion kits for Illumina sequencing libraries," in an attempt to smooth the transition to non-Illumina sequencer machines for interested institutions.[332]

The dynamics of the market for Illumina sequencer machines likely explain why and how the company established the near monopoly they enjoy today. Illumina derives significant profit from licensing patents and selling sequencer machines, which are routinely placed in: sequencing core facilities; individual labs; hospitals that run diagnostic genome sequencing for patients; and other sequencing companies that run whole-genome sequencing reactions to support academic research. In addition to the machines themselves, Illumina

---

327. Balasubramanian & Klenerman, *supra* note 185.

328. *See* Davies, *supra* note 186.

329. *Id.*

330. Andreas von Bubnoff, *Next-Generation Sequencing: The Race Is On*, 132 CELL 721 (2008).

331. *Id.* (explaining that most of the other competing companies have since filed for bankruptcy, switched out of the DNA sequencing space, or begun to pursue third-generation sequencing methods instead of NGS).

332. Michael Eisenstein, *Illumina faces short-read rivals*, 41 NATURE BIOTECHNOLOGY 3, 5 (2023) (quoting Shawn Baker, head of the genomics industry consultancy SanDiegOmics, saying that "[i]t's kind of a big deal, switching over" between sequencing platforms).

also sells ancillary technology, including: sample collection kits; DNA extraction reagents; library preparation kits; analysis software ("Basespace"); and other accessories that supplement sequencer machine functionality. Illumina also profits from service contracts with the various institutions that house Illumina sequencer machines.

Illumina's commercialization approach likely also explains their acquisition of Solexa. At the time of the merger, Illumina already dominated in the protein and RNA sequencing fields with their bead array technology, entirely independent of progress within the DNA sequencing market. The Solexa acquisition enabled Illumina to dominate across all macromolecule sequencing platforms, explaining the next decade of their success.

5. Aggressive Litigation

With the proliferation of genomics companies working across all three generations of sequencing technology, patent infringement suits between Illumina and other companies have surged in the past decade. A summary of many patent disputes between sequencing technology companies—including those without Illumina as a party—is available in an excellent 2012 review.[333] This Section briefly discusses Illumina's aggressive litigation strategy, involving allegations of infringement against many competitor sequencing machine companies.

Illumina began enforcing their patents via litigation as early as 2007, shortly after it merged with Solexa. For example, Illumina alleged that Applied Biosystems' sequencing-by-ligation SOLiD system infringed three of its patents directed to methods of chain elongation combined with solid support immobilization.[334] These patents came from Lynx Therapeutics (acquired in the Solexa merger in 2005). The inventor on the three Lynx patents was Stephen Macevicz, who previously worked as a patent attorney at Applied Biosystems.[335] Despite the inventor's association with the accused infringer, the Federal Circuit ruled in favor of Illumina.[336] Illumina continued to bring infringement suits, with a particularly strong peak occurring between 2009 and 2014.[337] Illumina has also initiated trade secret lawsuits against various companies.[338]

333. Holman, *supra* note 90.

334. U.S. Patent No. 5,750,341; U.S. Patent No. 5,969,119; U.S. Patent No. 6,306,597.

335. Holman, *supra* note 90, at 1056.

336. Applera Corp.-Applied Biosystems Grp. V. Illumina, Inc., 375 F. App'x 12 (Fed. Cir. 2010).

337. *See* Holman, *supra* note 90.

338. *See, e.g.*, Jonathan Wosen, *Illumina Sues Guardant Health, Saying Former Employees Stole Trade Secrets to Launch Liquid Biopsy Firm*, STAT BIOTECH (Mar. 17, 2022), https://

In addition to inter-genomics company litigation, Illumina has also been involved in multiple instances of FTC litigation, given its attempted acquisition of several other major companies in the sequencing space. Most notably, Illumina announced an initial merger agreement with Pacific Biosciences ("PacBio")—a third-generation sequencing company[339]—in 2018.[340] The FTC sued to block the merger, given Illumina's estimated ~90% share of the NGS market, over PacBio's estimated ~2–3% share.[341] Illumina and PacBio mutually agreed to terminate the merger in 2020.[342]

Similarly, in 2021, Illumina announced its intention to acquire Grail, a company that develops cancer tests.[343] While Grail did not work in the NGS market, the company's testing method relied on DNA sequencing.[344] The FTC sued to block this acquisition because "Illumina [was] the only provider of DNA sequencing that [was] a viable option" for the types of cancer tests developed by Grail.[345] In April 2023, the FTC ordered Illumina to divest Grail to block this vertical acquisition.[346]

---

www.statnews.com/2022/03/17/Illumina-sues-guardant-health-saying-former-employees-stole-trade-secrets-to-launch-liquid-biopsy-firm/.

339. *See* discussion *infra* Section V.C.

340. *Illumina to Acquire Pacific Biosciences for Approximately $1.2 Billion, Broadening Access to Long-Read Sequencing and Accelerating Scientific Discovery*, PACBIO (Nov. 1, 2018), https://www.pacb.com/press_releases/Illumina-to-acquire-pacific-biosciences-for-approximately-1-2-billion-broadening-access-to-long-read-sequencing-and-accelerating-scientific-discovery/.

341. Complaint, Illumina, Inc., FTC Matter No. 1910035 (Dec. 17, 2019); *Cancer Genomics Research*, ILLUMINA, https://www.illumina.com/areas-of-interest/cancer/research.html (last visited Sept. 23, 2022) (estimating that Illumina NGS and microarray technologies "account for ~90% of the world's sequence data").

342. *Illumina and Pacific Biosciences Announce Termination of Merger Agreement*, PACBIO (Jan. 2, 2020), https://www.pacb.com/press_releases/Illumina-and-pacific-biosciences-announce-termination-of-merger-agreement/.

343. Conor Hale, *Illumina to pay $8B to reacquire cancer blood test maker Grail, with all eyes on 2021*, FIERCE BIOTECH (Sept. 21, 2020), https://www.fiercebiotech.com/medtech/Illumina-to-pay-8b-to-reacquire-cancer-blood-test-maker-grail-all-eyes-2021.

344. Steve Lohr, *F.T.C. Orders Gene-Sequencing Company Illumina to Divest Acquisition*, N.Y. TIMES (Apr. 3, 2023), https://www.nytimes.com/2023/04/03/business/ftc-illumina-grail-divest.html.

345. Complaint, Illumina, Inc. and Grail, Inc., In the Matter of, FTC Matter No. 2010144 (Mar. 30, 2021), https://www.ftc.gov/legal-library/browse/cases-proceedings/201-0144-illumina-inc-grail-inc-matter.

346. *FTC Orders Illumina to Divest Cancer Detection Test Maker GRAIL to Protect Competition in Life-Saving Technology Market*, FED. TRADE COMM'N (Apr. 3, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-orders-illumina-divest-cancer-detection-test-maker-grail-protect-competition-life-saving.

V. STATE OF THE ART

Although Illumina has dominated the NGS market for over a decade, several alternative DNA sequencing approaches have proliferated on the sidelines. This Part discusses the modern state of DNA sequencing, focusing on the three generations: (1) the continued use of first-generation Sanger sequencing; (2) alternative strategies within the NGS approach; and (3) the development of long-read third-generation sequencing.

A. FIRST-GENERATION SEQUENCING AND TARGETED STUDIES

NGS provides advantages over first-generation sequencing techniques, including reduced speed and cost. Specifically, NGS increases the capacity to detect rare variants, achieves high coverage across entire genomes, and multiplexes several DNA libraries for parallelized analysis.[347] But, in many contexts, researchers prefer first-generation Sanger technology over NGS. For example, Illumina's NGS platform is less accurate for small-scale DNA sequencing for individual genes in low numbers.[348] For this purpose, Sanger sequencing remains the gold standard. And, even in diagnostic settings, NGS technologies have downsides. Many diseases are diagnosed based on the detection of chromosomal abnormalities in localized regions, integrating genetic and positional (*in situ*) data. In these diagnoses, targeted, low-throughput Sanger sequencing, often paired with assays such as fluorescence *in situ* hybridization, is more effective than whole-genome sequencing.[349]

As discussed in Section III.A, *supra*, Hood's research at Caltech catalyzed the early leap into automated Sanger sequencing. His company, Applied Biosystems (now Life Technologies), brought the first automated Sanger sequencing machine to market. Life Technologies, now a subsidiary of Thermo Fisher Scientific (Thermo),[350] remains a leader in the Sanger sequencing market.[351] Many core sequencing facilities around the world specialize in running Sanger sequencing reactions on Life Technologies and other automated sequencer machines. These Sanger sequencing methods support

347. Illumina on NGS vs Sanger, *supra* note 37.

348. *Id.* (describing that NGS is "less cost-effective" and "time-consuming" for researchers looking to "sequenc[e] low numbers of targets (1–20 targets)").

349. Behjati & Tarpey, *supra* note 76, at 236.

350. *Life Technologies*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/brands/life-technologies.html (last visited Nov. 24, 2022).

351. *See* Cook-Deegan & Heaney, *supra* note 239, at 404–5; *Instruments for Sanger Sequencing and Fragment Analysis by Capillary Electrophoresis*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-technology-accessories.html (last visited Oct. 24, 2022).

basic research in the life sciences, where small-scale confirmation of individual genomic regions is often an essential component of experiments.

B. ALTERNATIVE NGS METHODS

Many other companies have developed NGS technology with similar functionality as the Illumina platform. As discussed in Section II.B *supra*, there are four conserved steps in all NGS pipelines: library preparation, amplification, read generation, and data analysis. While most NGS methods carry out the first and final steps (library preparation and data analysis) similarly, the methods vary in the amplification and read generation steps.

As described in Section II.B.1.a), *supra*, Illumina carries out amplification with strands anchored to a solid support, leveraging bridge clustering and PCR amplification to generate sufficient substrate quantity for a strong fluorescent signal. The PCR step is unavoidable for NGS pipelines, but other companies replace solid support bridge clustering PCR with emulsion PCR.[352] Emulsion PCR maintains the physical immobilization of DNA strands by fixing template strands to beads, rather than a solid support array.[353] The PCR amplification step occurs on the beads, which are later immobilized by deposition on a different surface.[354] Multiple companies, including 454 Life Sciences and Applied Biosystems, adopted the emulsion PCR approach. 10x Genomics—although perceived by some to be a third-generation sequencing company—also relies on a bead-based immobilization step for their Chromium product line, feeding ultimately into an Illumina sequencing pipeline.[355] From a technical perspective, it remains unclear whether bridge PCR or emulsion PCR is more advantageous when integrated into NGS pipelines. While Illumina dominates the NGS market with a bridge PCR approach, emulsion PCR may achieve equally or even more better results—but the technique failed to take off due to Solexa's early market entry.

Companies in the NGS space have also explored modifications to the read generation step. Illumina uses SBS for read generation, with 3'-*O*-blocked nucleotides that emit fluorescent signals during sequencing. However, other SBS and non-SBS forms of read generation, distinct from the Illumina platform, may fit into the NGS protocol. For example, the first NGS machine on the market, 454 Life Sciences GS20, used a read generation method based on Melamede and Nyrén's early pyrosequencing work. As discussed in Section

352. *See* Metzker, *supra* note 42, at 32.

353. *Id.*

354. *Id.* (listing examples of emulsion PCR post-bead immobilization strategies, such as polyacrylamide gel on microscope sides, amino-coated glass surfaces, or PicoTiterPlate wells).

355. *Chromium Instrument Family*, 10X GENOMICS, https://www.10xgenomics.com/instruments/chromium-family (last visited Nov. 24, 2022).

II.B.1.c), *supra*, pyrosequencing is another form of SBS; it leverages the natural process of DNA strand elongation and the production of inorganic pyrophosphate to monitor luminescence as an alternative signal to the fluorescence of the Illumina platform. Roche Diagnostics acquired 454 Life Sciences, but has since been shut down.[356] In addition to their Sanger sequencing machines, Thermo also manufactures NGS systems that use yet another type of SBS, called Ion Torrent sequencing.[357] This approach, similar to pyrosequencing, leverages a natural byproduct released during DNA polymerization (hydrogen ions) to monitor pH as a "light-free" method of SBS.[358] Finally, an entirely non-SBS based form of NGS, called sequencing-by-ligation, substitutes DNA ligase for DNA polymerase. Applied Biosystems (later Life Technologies and, then, Thermo), previously marketed a version of this platform known as SOLiD (sequencing by oligonucleotide ligation and detection).[359]

C. THIRD-GENERATION SEQUENCING (TGS)

Despite completion of the HGP, for many years, the *complete* human genome sequence remained elusive. At least 5% of the human genome is littered with copy number variations, regions with atypical GC content, and repeats.[360] These genomic areas pose a unique technical challenge, even for NGS technologies, because repetitive regions are difficult to "read" accurately. So, 8% of the human genome was left unknown even after the HGP was completed, only to become accessible with the development of TGS technology.[361]

---

356. Mark Hollmer, *Roche to Close 454 Life Sciences as It Reduces Gene Sequencing Focus*, FIERCE BIOTECH (Oct. 17, 2013), https://www.fiercebiotech.com/medical-devices/roche-to-close-454-life-sciences-as-it-reduces-gene-sequencing-focus.

357. *Ion Torrent Next-Generation Sequencing Systems and Support*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-products-services/ion-torrent-next-generation-sequencing-systems-support.html (last visited Nov. 24, 2022).

358. *Ion Torrent Next-Generation Sequencing Technology*, THERMOFISHER SCI., https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html (last visited Nov. 24, 2022).

359. Anton Valouev et al., *A High-Resolution, Nucleosome Position Map of C. elegans Reveals a Lack of Universal Sequence-Dictated Positioning*, 18 GENOME RSCH. 1051 (2008).

360. Hood & Rowen, *supra* note 105, at 5.

361. Sarah Zhang, *The Human Genome Is—Finally!—Complete*, ATLANTIC (June 11, 2021), https://www.theatlantic.com/science/archive/2021/06/the-human-genome-is-finally-complete/619172/. The article reporting the truly complete sequence of the human genome was published in 2022. Nurk et al., *supra* note 35 (using a combination of technologies from PacBio, Oxford Nanopore, and Illumina, but predominantly long-read shotgun sequencing, to resolve the remaining 8% of the human genome sequence).

There are a few key differences between TGS and NGS technologies, although a precise definition of what constitutes a TGS platform remains in flux. A hallmark of NGS is the production of short reads (hundreds of base pairs in length).[362] TGS, on the other hand, typically produces reads over 10,000 base pairs in length.[363] Longer reads offer advantages in the context of genome assembly, where piecing together an organism's entire genome benefits from large regions of overlap between individual reads.[364] This difference highlights the complexity of assembly and data analysis at the end of the sequencing pipeline: short-read technologies only generate complete human genome sequences when they have a reference sequence for assembly; long-read technologies assemble genome sequences *de novo*.[365]

Another difference is that TGS methods omit the PCR amplification step of NGS. NGS relies on PCR amplification to generate an abundance of template DNA, so that the resulting fluorescent signal from SBS read generation is strong enough for detection. However, PCR is "cumbersome to implement" in the context of high-throughput DNA sequencing.[366] The PCR amplification step makes NGS vulnerable to erroneous introduction of mutations (which show up as false positive variants) and amplification bias (which arbitrarily alters the relative abundance of sequence fragments).[367] For highly quantitative applications, such as rare variant analysis, these PCR-inherent qualities can pose a serious issue.[368]

Unlike Illumina's dominance in the NGS space, there are several central players with currently or previously successful TGS platforms, including Helicos Biosciences (Helicos), PacBio, and Oxford Nanopore Technologies (Oxford). Researchers tend to prefer these platforms over NGS for *de novo* genome assembly, which benefits from longer reads.[369] All, however, are currently hindered by much higher error rates than other types of sequencing technologies. The biochemical basis of TGS (i.e., long DNA strands) is inherently less stable than the shorter strands used by NGS technology; as

362. Jade L. L. Teng et al., *PacBio but not Illumina Technology can Achieve Fast, Accurate and Complete Closure of the High GC, Complex Burkholderia pseudomallei Two-Chromosome Genome*, 8 FRONTIERS MICROBIOLOGY 1, 2 (2017).

363. Alice McCarthy, *Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology*, 17 CHEMISTRY & BIOLOGY INNOVATIONS 675, 675–76 (2010).

364. Teng et al., *supra* note 362, at 11 (noting that "long reads have greatly enhanced the accuracy of genome assembly").

365. Hood & Rowen, *supra* note 105, at 6.

366. Metzker, *supra* note 42, at 32.

367. *Id.*

368. *See id.*

369. *See* Metzker, *supra* note 42, at 37.

reads increase in length, the quality of sequencing data gradually deteriorates.[370] This is unacceptable for many of the more sensitive applications of NGS, but permissible for more crude purposes, such as structural variant calling.[371]

Several companies pursuing TGS instead use a *single-molecule*, amplification-free sequencing approach. Recall that the Solexa team was initially working with a single-molecule platform, before licensing the bridge PCR cluster technology from Manteia—this pushed them from single-molecule sequencing into the conventional NGS amplification approach. Helicos stuck with this single-molecule strategy and immobilized strands to a solid support without amplifying them.[372] And for Helicos, this worked—their platform was the first to successfully implement the single-molecule strategy to sequence DNA, building on preliminary research conducted by Stephen Quake at Caltech. Helicos has since filed for bankruptcy, having stepped out of the DNA sequencing space in 2010.[373] PacBio adopted a different approach, using polymerase molecules (themselves immobilized on a solid support) to immobilize single strands of DNA.[374] In the PacBio process, a single molecule passes through the polymerase as sequencing occurs in real-time. PacBio remains a leader in the TGS space, with Illumina recently (unsuccessfully) attempting to acquire the company due to an FTC complaint.[375] And Oxford uses yet another approach: it distinctively achieves template immobilization using biological protein nanopores and sequences single DNA molecules as they pass through the pore.[376] With the transition away from a solid support array, the Oxford platform is miniaturized and portable. Notably, many of the original Solexa team members, including Milton, Brown, and McCooke, now work at Oxford.

Another key difference between TGS and NGS technologies is in the read generation step. Illumina's NGS uses reversible terminator chemistry as part

370. Sara Goodwin et al., *Coming of Age: Ten Years of Next-Generation Sequencing Technologies*, 17 NATURE REVS. GENETICS 333 (2016).

371. Structural variants are genomic changes involving regions greater than 50 base pairs in length. TGS methods have been shown to outperform NGS technologies in structural variant calling. *See* Jason D. Merker et al., *Long-Read Genome Sequencing Identifies Causal Structural Variation in a Mendelian Disease*, 20 GENETICS MED. 159 (2018); Fritz J. Sedlazeck et al., *Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing*, 15 NATURE METHODS 461 (2018).

372. Metzker, *supra* note 42, at 33.

373. Helicos Biosciences Corp. (Form 8-K) (Nov. 15, 2012), https://web.archive.org/web/20121121065028/http://biz.yahoo.com/e/121115/hlcs8-k.html.

374. Metzker, *supra* note 42, at 33.

375. *See* discussion *supra* Section IV.B.5.

376. Daniel Branton et al., *The Potential and Challenges of Nanopore Sequencing*, 26 NATURE BIOTECHNOLOGY 1146 (2008).

of SBS, with 3'-*O*-blocked nucleotides that both transiently pause chain elongation and fluoresce as an approximation of nucleotide identity. The Illumina nucleotides contain a blocking group attached to the ribose sugar, and a fluorescent label attached to the nucleobase. Helicos, PacBio, and Oxford all adopted different nucleotide chemistry for this step of their TGS platforms. Helicos used 3'-*O-unblocked* nucleotides, with both a blocking group and a fluorescent label attached to the nucleobase.[377] PacBio uses yet another distinct nucleotide type, which omits the blocking group and includes a fluorescent label alone, attached instead to the phosphate groups.[378] Finally, Oxford distances itself entirely from the SBS chemistry concept; it monitors changes in electric current as single DNA strands pass through the nanopores, rather than monitoring fluorescent signals.[379]

VI. CONCLUSION

NGS technology has been one of the most important developments in molecular biology since the 1970s. NGS platforms revolutionized the ways in which scientists and clinicians approached the analysis and treatment of disease. The technology continues to break new ground in bringing research from bench to bedside. This Article focused on the Illumina NGS platform as the dominant technology of the DNA sequencing market for over a decade.

As of 2023, Illumina maintains control over an estimated 80% of the sequencing market.[380] But, given the recent—and imminent—expiry of some of Illumina's major SBS patents, this dominance may falter soon.[381] Illumina also increasingly faces competition from TGS companies. And the price of sequencing a human genome continues to fall below even the level HGP researchers targeted. Certainly, however, Illumina's work has laid a remarkable base for future development in DNA sequencing.

The history of Illumina's success took place over several decades, across multiple countries and institutions. It occurred in two distinct stages—one dedicated to scientific discovery, the other to business development. Many different sources of motivation, features of scientific research, and characteristics of intellectual property protection all came together to propel the Illumina NGS platform forward. This development story suggests that an

377. Timothy D. Harris et al., *Single-Molecule DNA Sequencing of a Viral Genome*, 320 SCI. 106 (2008); Metzker, *supra* note 42, at 34–35.

378. Paul M. Lundquist et al., *Parallel Confocal Detection of Single Molecules in Real Time*, 33 OPTICS LETTERS 1026 (2008); Metzker, *supra* note 42, at 35.

379. Branton et al., *supra* note 376.

380. Eisenstein, *supra* note 332, at 3.

381. *Id.*

initial era of scientific curiosity, altruism, public funding sources, academic recognition, and serendipity established an initial foundation from which the Solexa team grew. Then, as research and development costs increased, a secondary era driven by private funding sources, a rigorous patent portfolio, well-timed licensing, dedication to commercialization potential, and aggressive litigation brought the technology to market. For this platform technology—and perhaps for many other areas of science—it does seem true that "so much progress depends on the interplay of techniques, discoveries and new ideas, probably in that order of decreasing importance."[382]

382. Sydney Brenner, Symposium Talk at the Friedrich Miescher Institute, Basel, Switzerland (Mar. 20, 1980), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139404/.