Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Nat Biotechnol. Author manuscript; available in PMC 2024 April 07.

Published in final edited form as:
Nat Biotechnol. 2023 December ; 41(12): 1709–1715. doi:10.1038/s41587-023-01986-3.

Genesis of next-generation sequencing

Raphaël Rodriguez[1], **Yamuna Krishnan**[2,3,4]

[1]Institut Curie, CNRS, INSERM, PSL Research University, Equipe Labellisée Ligue Contre le Cancer, Paris, France

[2]Department of Chemistry, The University of Chicago, Chicago.

[3]The Neuroscience Institute, The University of Chicago, Chicago

[4]Institute of Biophysical Dynamics, The University of Chicago, Chicago

Abstract

The first large genome that was fully sequenced by next-generation sequencing (NGS) was that of a bacteriophage. For NGS to grow from rudimentary proof-of-concept and a few bench-top experiments to being both affordable, and useful on a practical timescale to impact humanity, it needed and followed a non-traditional scientific path. The key scientific advances were not disseminated in the standard fora. The developments outlined in this Perspective are reconstructed from past conversations with some of the inventors of NGS when they were exploring chemistries to sequence even just two bases. Although not privy then to those specific chemistries that finally worked, we suggest the potential problems encountered at each stage. Now that we know the chemistries that work, for the benefit of future generations of scientists, we chart how NGS was developed and thereby illustrate the rationale underlying the enabling chemistry that has impacted biology and indeed society in unprecedented ways.

Introduction

The first map of a human genome was established using Sanger Sequencing and was reported in 2001. It took thirteen years, a transcontinental effort and an excess of $0.5 billion to accomplish[1,2]. Today, using Illumina sequencing, referred to as NGS in this Perspective, a human genome can be sequenced within 48 hours for under $200 and with much higher accuracy. This speed and economy of effort has been the bedrock of our response to the COVID-19 pandemic. While large scale PCR and antibody testing helped restrain the SARS-CoV-2 outbreak, NGS was used decipher the viral genome sequence and identify variants of concern, which concurrently informed the design of vaccines[3]. Furthermore, NGS helped pinpoint pandemic hotspots, by tracking the mutating SARS-CoV-2 as the pandemic wore on.

Sequencing by electrophoresis or other size-based separation of nucleic acid fragments are limiting both in cost and time efficiency. Hence, in order to sequence large genomes, a different paradigm was needed. In 1990, Tsien and co-workers outlined a series of

Correspondence should be addressed to raphael.rodriguez@curie.fr and yamuna@uchicago.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

strategies in a patent that conceptualized sequencing-by-synthesis (SBS) using automated instruments[4]. Notably they hypothesized that a the 3′-OH of dNTPs could be protected with a reporter group to enable the stepwise incorporation and detection of dNTPs by a polymerase, base-by-base, using an immobilized DNA template. Though this patent was discontinued shortly thereafter, others pursued their quests towards developing effective SBS technologies. In what we consider the precursor to their idea, Balasubramanian and co-workers had investigated fluorescence energy transfer between the Klenow fragment of the DNA polymerase and a series of DNA primer templates, which positionally mapped nucleobases along the DNA helix at the single molecule level[5]. The success of this experiment led to their idea that single molecule fluorescence and microscopy might provide that alternative paradigm for DNA sequencing, if the DNA was surface-immobilized and if every incoming nucleotide had a unique fluorophore, so that one could tell the bases apart as they were being incorporated. At the time, there were several distinct challenges that each needed solving to make this idea work. These included identifying an appropriate protecting group at the 3′-OH of the deoxyribose to enable the incorporation of nucleotides one at a time, tethering fluorophores on each of the four nucleotides such that it would not disrupt nucleic acid synthesis, engineering a polymerase that would tolerate these modifications, enable fluorescence detection and reversal of the terminating group, developing appropriate surface chemistry to retain the strand being sequenced over hundreds of cycles of nucleotide incorporation and finally inventing a system to circumvent low fluorescence signal-to-noise ratio and to rectify potential errors due to dNTP misincorporation by the polymerase. We describe below how each of these problems were solved, leading to the chemistry as it stands today. NGS turned out to be much faster, more accurate and cheaper compared to other methods because it moved away from capillaries and gel electrophoresis, enabling a massively parallel sequencing approach. First reported In 1993, pyrosequencing was an early version of SBS based on inorganic phosphate detection by luminometry. From 2005–2008, sequencing technologies distinct from NGS have also materialized. Third-generation (‘post light’) sequencing has emerged as complementary alternatives to NGS. Notably, Pacific Biosciences and Oxford Nanopore Technologies developed single molecule real time sequencing (SMRT) and pore sequencing, respectively. Conceptually different from NGS, these methods are based on the electronic analysis of molecules and are true single molecule analysis technologies. While these techniques may be more expensive, lower throughput or less accurate than NGS, they enable long sequencing reads more effectively. These and other NGS technologies distinct from Illumina sequencing are described in greater detail elsewhere[6–9].

Developing an enabling reversible terminator chemistry

Sanger Sequencing is based on the synthesis of truncated DNA fragments in solution from multiple DNA templates, prematurely terminating elongating strands via the stochastic incorporation of labeled dNTPs lacking a 3′-OH (ddNTP)[10]. This is the earliest form of the so-called ‘terminator chemistry’, where truncated synthetic DNA strands are size-separated by gel electrophoresis from which the original sequence is reconstructed (Fig. 1). In the new sequencing paradigm, short DNA reads would be obtained from fragments of genomic DNA, immobilized on a solid surface, and would serve as templates for a polymerase. Each

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of these fragments would be replicated (synthesized) in parallel and the elongating strands simultaneously scanned by fluorescence imaging. To this end, it was envisioned that the 3′-OH of dNTPs could be exploited to control the stepwise incorporation of each nucleotide.

Back in the 90s, several groups were seeking to develop new ways to establish an efficient SBS method and to that end, distinct chemistries were explored. Many creative strategies had shortcomings. The chemistry that finally worked, did so well enough to satisfy all the key requirements, which ultimately enabled rapid DNA sequencing at scale and low cost. Early work focused on the effect of fluorophores at the 3′-OH of dNTPs, designed to serve a dual purpose: act as a protecting group for terminator chemistry and as a fluorescent marker to identify the incorporated nucleotide[11]. While conceptually attractive, it was later found that the 3′-OH points towards amino acid residues lining the enzymatic pocket of DNA polymerases[12]. Hence, bulky moieties like large, fluorescent protecting groups at this position prevent fast incorporation of dNTPs with the required fidelity. While fluorophores can be tethered to the 3′-OH via various types of linkers[11], the low deprotection rate would make this strategy, involving several rounds of cleavage, ill-suited to sequence large genomes. For example, a time of 2 hours per deprotection would require 25 days to sequence a DNA fragment just 300 bases long. Hence Balasubramanian and co-workers quickly realized that fluorescently labeled dNTPs at the 3′-OH was not viable.

Indeed, others also explored the use of protecting groups at the 3′-OH for terminator chemistry[13]. For example, the use of 2-nitro-phenol was appealing because of the fast cleavage rate upon exposure to UV light, leaving the 3′-OH available for the next round of synthesis. However, the use of UVs was not ideal given that DNA is prone to photodamage. Repetitive cycles of irradiation would chemically alter the DNA being sequenced.

Alternatively, it was found that tethering a bulky fluorophore on the nucleobases fortuitously acted as a terminating agent just by virtue of steric hindrance, thus obviating the need to protect the 3′-OH[14]. Chemical removal of the fluorophore following its incorporation on the strand would then allow the next nucleotide to be incorporated. However, using Alexa Fluor 594 to label all four dNTPs – a single color readout - imposed an experimental design where the DNA template needed to be incubated with each of the four dNTPs, one at a time, with the hope that only the correct nucleotide would be incorporated. This however would substantially increase the number of steps of the sequencing procedure and it is also prone to misincorporation of dNTPs as the chemistry itself is non-competitive. Indeed, in absence of correct dNTP to be incorporated, polymerases can incorporate those available instead. This meant that the reversible terminator chemistry and fluorescence detection had to be uncoupled from one another.

A viable strategy would require a protecting group for the 3′-OH that was small enough to be a competent substrate for DNA polymerases, highly stable under the aqueous conditions used for DNA synthesis, cleavable in quantitative yields with a fast turnover, and all of this under reaction conditions that preserved the integrity of DNA. It was found that an azide-based protecting group, a non-naturally occurring building block, could cap the 3′-OH. Further, its removal could be triggered by a water-soluble phosphine via a Staudinger reaction[15–17]. This bio-orthogonal strategy enabled accurate base-by-base incorporation of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dNTPs using four differently labeled fluorescent dNTPs, bearing fluorophores on the base itself as opposed to the sugar moiety (Fig. 2).

This milestone was crucial for fluorescence imaging of the elongating DNA strand – its rapidity and efficacy were foundational to sequencing large genomes on practical timescales. For over a century, chemists have attempted to build complex, naturally occurring molecules by mimicking reactions and using substrates found in Nature[18,19], with the idea that biomimetic strategies are more elegant, more straightforward and have greater 'atom economy' compared to those that rely on protecting groups[20–22]. Indeed, DNA polymerases synthesize DNA without protecting groups. In the context of NGS, however, it is the use of a protecting group and a chemistry not found in Nature that enables the controlled, stepwise incorporation of dNTPs. NGS is an example of how chemistry can provide new and powerful tools for biology and medicine, through the design of unnatural compounds that rival natural ones in terms of reactivity and selectivity.

Developing fluorescently-labelled dNTPs for imaging

Along with co-workers at Solexa, Balasubramanian and Klenerman deduced that a working solid-surface-based NGS technology required four 'reversible terminator' dNTPs designed such that each nucleobase displayed a distinct cleavable fluorophore[23]. This would reveal the DNA code as the polymerase elongates the DNA strand complementary to the template. Previous work had shown for 'non-reversible terminator' chemistry that a reporter could be directly linked to a nucleobase without affecting Watson-Crick-Franklin base pairing, as long as it jutted out of the outer edge of the nucleobases, an area referred to as the Hoogsteen face[24,25].

Potential limitations of this strategy included the rate of polymerase elongation, which could be affected by large fluorophores as well as misincorporation of unnatural dNTPs. Further, it was desirable for the dNTPs to have spectrally distinct fluorescence emissions. The team at Solexa successfully designed four dNTPs, each displaying a distinct fluorophore tethered by a cleavable linker that can be removed after each round of dNTP incorporation. The fluorescent guanine nucleobase, in particular, was difficult to develop, as guanine tends to quench fluorophores due to single electron transfer resulting from its low redox potential. So, the fluorescent GTP monomer was designed to have a different chemistry to neutralize the single-electron transfer between the fluorophore and the nucleobase. Importantly, all linkers were designed such that after the fluorophore was removed, the bases displayed residual appendages that were non-bulky and chemically bio-orthogonal, thereby minimizing perturbations of the resulting double-stranded DNA[17].

Engineering a suitable DNA polymerase for SBS

SBS works through reversible terminator chemistry, which enables the incorporation of dNTPs base by base. Natural polymerases have evolved to incorporate natural dNTPs without pausing after each dNTP incorporation and with high speed. Having done his postdoctoral work with S. Benkovic at Penn State University, who was studying the mechanism of nucleotide incorporation by various polymerases at the time,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Balasubramanian knew from the start that solving the sequencing problem comprised at least two parts. One part was getting the nucleotide chemistry right, and the other was developing a polymerase to accommodate that new chemistry. This was all happening at a time that preceded the atomic-level structural elucidation of nucleotide incorporation. Thus, from mechanistic insights arising at that time, one had only a hazy map of the major residues and dynamics within the active site of the polymerase. Thus, the rationale presented below predated any knowledge of the structure of the active site and how the polymerase worked at the atomic level.

The polymerase developed at Solexa had to tolerate the 3′-OH protecting group better than wild-type polymerases and yet had to exhibit comparable kinetics of incorporation. Previous work had shown that polymerases could indeed tolerate modifications at the 3′-OH, but the incorporation rates were slow and therefore they needed higher concentrations of the modified nucleotides[4]. Solexa's polymerase needed to be as accurate with these modified bases, as the natural one was for natural dNTPs. Thus, the chemistry was designed to have each of the four dNTPs compete with each other for the enzyme active site. The accuracy stemmed from having the right dNTP in addition to the wrong ones during the sequencing reaction to reduce misincorporation. There are a number of steps that guide accuracy when a polymerase incorporates a base onto a growing strand.[26]

One of these occurs at the level of dNTP binding. Repeated on and off binding allows the dNTP to sample the base on the template strand. Next, there is a conformational change that locks all the catalytic, reactive, functional groups, which aligns the 3′-O, the αP atom of the dNTP and the oxygen bridging to the β-γ phosphates. This allows an in-line nucleophilic attack of the 3′-OH that leads to an inversion at the αP (Fig. 3). When it all locks in, a mismatch between the incoming nucleotide and the template nucleobase introduces a kinetic barrier that forces the enzyme to go backwards and bind a different dNTP. Thus, they had to engineer and screen different polymerases to optimize both steps, but with the modified dNTPs. To this end, by saturation mutagenesis on three key amino acids on different polymerases, they selected a suitable enzyme using chemically modified dNTPs.

The other feature they needed to tune the polymerase for was to get it to fall off the DNA strand more easily after incorporating the correct dNTP, unlike natural polymerases which have evolved to do the opposite to preserve processivity. In SBS, no incorporation can take place after dNTP incorporation because the 3′-OH is protected. Even so, if polymerases stay bound to elongating DNA strands, it would likely impede the incorporation of dNTPs in all the strands of a given cluster due to steric crowding. Since the rate-limiting step for a single dNTP incorporation is the unbinding of the polymerase to the template, the polymerase had to be re-engineered to retain accuracy and speed of dNTP incorporation. However, its processivity had to be altered. They realized that they needed to increase the off-rate without affecting anything else on the polymerase such that complete incorporation of the dNTP could be achieved on each elongating strand within a colony prior fluorescence acquisition and removal of the protecting group. Therefore, they mutated out two or three positively charged amino acids on the polymerase near the binding site, but away from the catalytic site. This increased the dissociation constant (K_d), by virtue of a higher off-rate (k_{off}), which made the sequencing method viable[27–29].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Developing an enabling reversible-terminator chemistry

To sequence large genomes on practical timescales, one needed to sequence millions of smaller but distinct fragments in parallel, then piece the information from these individual fragments together correctly and thereby reconstruct the sequence of the whole genome. This could only be achieved on a solid support, where each fragment could be investigated in spatial isolation. This automatically ruled out solution-based approaches. As previously posited by Tsien[4], each DNA template strand had to be immobilized on a solid surface so that one could pinpoint which nucleotide was incorporated on which strand using fluorescence microscopy. Since SBS involved repeated cycles of dNTP incorporation to sequence each of the immobilized short reads, Solexa needed to develop surfaces and DNA-attachment chemistry that could withstand repeated exposure to the sequencing reaction conditions. Otherwise, DNA strands that were being sequenced would be lost or leached off during the process. The surface also had to be passive to prevent background fluorescence that could interfere with the fluorescence of incorporated dNTPs. A key challenge was also to prevent the non-specific binding of fluorescent dNTPs to the surface, since these were needed at micromolar concentrations for the polymerase to work effectively, and only a single copy of DNA template was replicated with base-by-base dNTP incorporation in the original single molecule version. This non-specific background needed to be drastically reduced by surface engineering so as to detect genuine nucleotide incorporation events. While this was partly achieved, the background progressively increased over large cycle numbers. Some of the chemistries that were explored early on included silane glass coating, activated benzene triacetic acid loading then amino-DNA grafting. Polyacrylamide surface chemistry offered more stability and a higher signal-to-noise ratio[30]. Illumina sequencing platforms have evolved tremendously over the past two decades and the surface chemistry used in the latest apparatus has, to the best of our knowledge, not been reported. Furthermore, DNA polymerase itself has inherent errors of misincorporation, albeit low ones. Thus, another innovation had to be integrated at this stage to circumvent these limitations to enable NGS to work at scale.

Enabling parallelization for speed and accuracy

An obvious challenge to the original plan for single molecule sequencing was the requirement to detect fluorophores with extreme sensitivity. At this point, it was clear that sequencing unique copies of DNA fragments would invariably yield low signal-to-noise ratio. Solexa therefore needed to deviate from the strategy as initially conceived and incorporated a technology developed at Manteia Predictive Medicine by Mayer and co-workers[31]. This technology, termed molecular clustering, was designed to locally amplify a single, surface-immobilized DNA fragment into a colony of hundreds of identical copies that were also surface-immobilized around the original DNA strand. Thus, going from single molecule fluorescence to “few-molecules fluorescence” had two major advantages. It improved the signal-to-noise ratio substantially. Importantly, it also compensated for stochastic misincorporation events because it averaged out the signal from misincorporation in a single strand over many simultaneous correct incorporations on the remaining strands within a given colony.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In molecular clustering, specific primers are attached on both extremities of every genomic DNA fragment to be sequenced. These modified DNA fragments are then hybridized to the surface, which displays DNA sequences complementary to these primers. Thus, amplification with a DNA polymerase results in the local production of multiple copies of forward and reverse strands, which is facilitated by the amplified sequence forming a bridge with another adjacent surface-immobilized primer (Fig. 4).

This bridge amplification strategy, so-called DNA-bridge chemistry, required many iterations to be optimized and successfully incorporated into the final workflow. For example, the complementary DNA strands rehybridize quickly, which competes with their desired hybridization to the surface-immobilized primers. This was solved by Mayer, implementing an isothermal amplification on the surface[32]. Further, the uneven fluorescence background of dyes hindered accurate signal quantification from a given DNA colony, requiring the use of a locally probabilistic cellular automaton approach. Finally, the workflow was designed to be able to image large areas of a slide with single molecule sensitivity, ensuring that each cluster remained aligned across every new cycle of imaging. The optical system was fixed, whereas the stage with flow-cell onto which the DNA to be sequenced is mobile, so that large areas could be imaged and the sequencing could be massively parallelized (Fig. 5)[17,33]. Typically, multiplying fragments enabled acquisition of signal a thousand time faster.

Conclusions

While the pandemic emphasized the value of genome sequencing at scale, speed and economy, NGS had already initiated a quiet revolution across the spectrum of life sciences, where the ease of DNA sequencing provides the means to interrogate biology and to solve medical problems in unprecedented ways. NGS has enabled the identification of risk genes for diseases such as autism where the underlying genetics was unclear[34]. It illuminated chromatin biology, providing insights into the genomic locations of chromatin marks[35], transcription factor binding loci[36] and genomic sites of action of small molecules[37,38]. It has been used as a molecular counter to quantify RNA transcripts in tissues with spatial resolution[39,40]. NGS impacted clinical medicine by revealing disease-causing mutations[34], providing critical information that guided therapeutic strategies and enabling non-invasive circulating cell-free DNA analysis as diagnostic markers[41]. More recently, NGS made *in vivo* genome editing possible by providing sequence maps of genomic loci that can be altered and to validate selective editing[42]. In addition, from analyses of big data, NGS now allows to establish ancestry, solve cold cases and exonerate innocents on death-row[40]. Finally, NGS is reshaping our understanding of genome evolution, providing insights into future evolutionary trends across species.

The first genome sequenced using Solexa's approach was that of the phiX-174 bacteriophage in 2005, which demonstrated the advancement over Sanger Sequencing by generating substantially more data in a single run. Solexa launched the Genome Analyzer 1st Generation (GA1), its first commercial device shortly thereafter. It was then acquired by Illumina, whereafter hundreds of scientists further improved the technology. The first human genomes were then sequenced using Solexa/Illumina NGS technology[17,33]. It was

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mainly the result of a team effort and collective intelligence that led to a SBS strategy made possible by the key development of reversible terminator chemistry. It took several years to define the key steps to achieve such a technology, exploring the many possible routes forward and finally identifying one that works. It required scientists with a multidisciplinary expertise spanning structural biology, biochemistry, chemical synthesis, surface chemistry, optical physics, computing and managerial skills. Here, we described the key elements that went into the development of a technology that turned a pipe dream into reality. From the breadth of its impact on biology, medicine and public health, NGS illustrates how central Chemistry is to Science. The development of NGS is the result of not only thinking outside the box, risk taking and many failed experiments, but the realization upfront of what would make it a successful technology and the capacity to identify key problems to be solved along the way.

Acknowledgment

We thank all those whose intellectual and practical contributions have collectively enabled the development of NGS. The intellectual property underlying NGS is protected by a series of patents not discussed in this article. We thank J. Szostak, V. Ramakrishnan and J.-M. Lehn for fruitful discussions and our colleagues at the University of Chicago and the Institut Curie for their constructive comments.

R.R. was supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 647973), Foundation Charles Defforey-Institut de France and Ligue Contre le Cancer. Y.K. is supported by NIH grant 1DP1GM149751-01 and the Ono Pharma Foundation.

References

1. Initial sequencing and analysis of the human genome. International Human Genome Consortium. Nature 409, 860–921 (2001) [PubMed: 11237011]
2. The sequence of the human genome. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ et al. Science 291, 1304–1351 (2001) [PubMed: 11181995]
3. A novel coronavirus from patients with pneumonia in china, 2019. Zhu N, Zhang D, Wang W, Li X, Yang B et al. N. Engl. J. Med 382, 727–733 (2020) [PubMed: 31978945]
4. DNA sequencing. Roger RY, Ross P, Fahnestock M, Johnston A. WO9106678A1
5. Use of fluorescence resonance energy transfer to investigate the conformation of DNA substrates bound to the Klenow fragment. Furey WS, Joyce CM, Osborne MA, Klenerman D, Peliska JA, Balasubramanian S. Biochemistry 37, 2979–2990 (1998) [PubMed: 9485450]
6. Gupta N and Verma VK (2019). Next-generation sequencing and its application: empowering in public health beyond reality. In: Arora P (eds) Microbial Technology for the Welfare of Society. Microorganisms for Sustainability, vol 17. Springer, Singapore. 10.1007/978-981-13-8844-6_15
7. Milestones in genomic sequencing. Koch L, Potenski C, Trenkmann M. Nature Milestones. S3 (2021)
8. Sequencing – the next generation. Willson J. Nature Milestones. S7 (2021)
9. The dawn of personal genomes. Burgess D. Nature Milestones. S9 (2021)
10. DNA sequencing with chain-terminating inhibitors. Sanger F, Nicklen F, Coulson AR. Proc. Natl. Acad. Sci. U. S.A 74, 5463–5467 (1977) [PubMed: 271968]
11. DNA polymerase fluorescent substrates with reversible 3′-tags. Canard B, Sarfati RS. Gene 148, 1–6 (1994) [PubMed: 7523248]
12. Syntheses of nucleosides designed for combinatorial DNA sequencing. Welch MB, Martinez CI, Zhang AJ, Jin S, Gibbs R, Burgess K. Chem. Eur. J 5, 951–960 (1999)
13. Termination of DNA synthesis by novel 3′-modified deoxyribonucleoside 5′-triphosphates. Metzker ML, Raghavachari R, Richards S, Jacutin SE, Civitello A, Burgess K, Gibbs RA. Nucleic Acids Res. 22, 4259–4267 (1994) [PubMed: 7937154]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

14. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Turcatti G, Romieu A, Fedurco M, Tairi A-P. Nucleic Acids Res. 36, e25 (2008) [PubMed: 18263613]
15. Über neue organische phosphorverbindungen III. Phosphinmethylenderivate und phosphinimine. Staudinger H and Meyer J. Helv. Chim. Acta 2, 635 (1919)
16. Sequencing nucleic acids: from chemistry to medicine. Balasubramanian Shankar. Chem. Commun 47, 7281–7286 (2011)
17. Accurate whole human genome sequencing using reversible terminator chemistry. Bentley DR et al. Nature 456, 53–59 (2008) [PubMed: 18987734]
18. A synthesis of tropinone. Robinson R. J. Chem. Soc. Perkin Trans I, 762–768 (1917)
19. Total synthesis of cyercene A and the biomimetic synthesis of (±)-9,10-deoxytridachione and (±)-ocellapyrone Rodriguez AR, Adlington RM, Eade SJ, Walter MW, Baldwin JE, Moses JE. Tetrahedron 63, 4500–4509 (2007)
20. Greene's protective groups in organic synthesis, fourth edition. Wuts PGM, Greene TW (2006), John Wiley & Sons, Inc.
21. Total synthesis of marine natural products without using protecting groups. Baran PS, Maimone TJ, Richter JM. Nature 446, 404–408 (2007) [PubMed: 17377577]
22. The atom economy—a search for synthetic efficiency. Trost BM. Science 254, 1471–1477 (1991) [PubMed: 1962206]
23. Labelled nucleotides. Barnes C, Balasubramanian S, Liu X, Swerdlow H, Milton J. US 7,057,026 B2
24. Synthesis of fluorescent or biotinylated nucleoside compounds. Sarfati SR, Pochet S, Guerreiro C, Namane A, Huynh-Dinh T, Igolen J. Tetrahedron 43, 3491–3497 (1987)
25. New dye-labeled terminators for improved DNA sequencing patterns Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM. Nucleic Acids Res. 25, 4500–4504 (1997) [PubMed: 9358158]
26. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Kiefer JR, Mao C, Braman JC, Beese LS. Nature 391, 304–307 (1998) [PubMed: 9440698]
27. Modified polymerases for improved incorporation of nucleotide analogues. Smith GP, Bailey DMD, Sanches RM, Swerdlow H, Earnshaw DJ. WO 2005/024010 A1
28. Improved polymerases. Ost TWB, Smith GP, Balasubramanian S, Rigatti R, Sanches RM. WO 2006/120433 A1
29. Modified polymerases for improved incorporation of nucleotide analogues. Smith GP, Bailey DMD, Sanches-Kuiper RM, Swerdlow H, Earnshaw DJ. US 8,852, 910 B2
30. Modified molecular arrays. Smith M, Brennan E, Sabot A, Rasalonjatovo IMJ, Sohna Sohna J-E, Horgan AM, Swerdlow HP. WO 2005/065814 A1
31. Method of nucleic acid amplification by extension of immobilized primers. Kawashima E, Farinelli L, Mayer P. WO 9844151
32. Isothermal amplification of nucleic acids on a solid support. Mayer P. WO 02/46456
33. The diploid genome sequence of an Asian individual. Wang J et al. Nature 456, 60–66 (2008) [PubMed: 18987735]
34. The next-generation sequencing revolution and its impact on genomics. Koboldt DC, Meltz Steinberg K, Larson DE, Wilson RK, Mardis ER. Cell 155, 27–38 (2013) [PubMed: 24074859]
35. High-resolution profiling of histone methylations in the human genome. Barski A et al. Cell 129, 823–837 (2007) [PubMed: 17512414]
36. Genome-wide mapping of in vivo protein-DNA interactions. Johnson DS, Mortazavi A, Myers RM, Wold B. Science 316, 1497–1502 (2007) [PubMed: 17540862]
37. Small molecule-induced DNA damage identifies alternative DNA structures in human genes. Rodriguez R, et al. Nat. Chem. Biol 8, 301–310 (2012) [PubMed: 22306580]
38. Unravelling the genomic targets of small molecules using high-throughput sequencing. Rodriguez R, Miller KM Nat. Rev. Genet 15, 783–796 (2014) [PubMed: 25311424]
39. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Nat. Methods 5, 621–628 (2008) [PubMed: 18516045]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

40. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Ståhl PL et al. Science 353, 78–82 (2016) [PubMed: 27365449]
41. DNA sequencing at 40: past, present and future. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. Nature 550, 346–353 (2017)
42. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. Newby GA et al. Nature 595, 295–302 (2021) [PubMed: 34079130]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

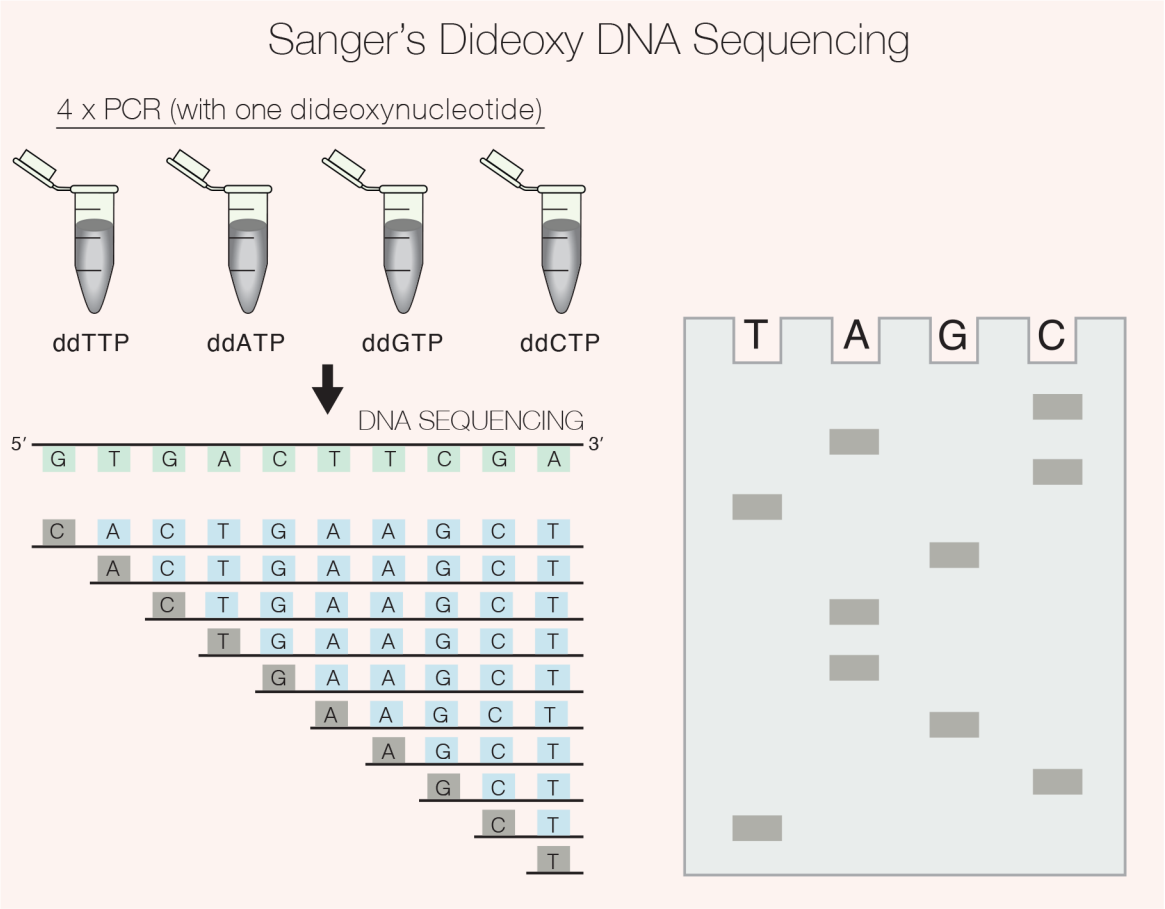

Figure 1.

Schematic illustration of Sanger Sequencing. DNA strands are separated based on size, which is defined by where into the chain a ddNTP is incorporated. This reveals the identity of the nucleobase at any position, as incorporation of a ddNTP terminates the growing chain at that position.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

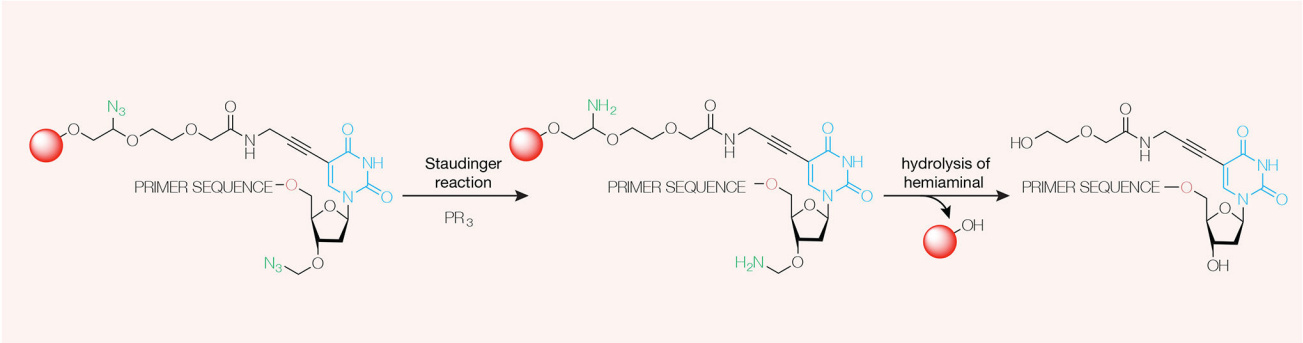

Figure 2.

Scheme of reversible terminator chemistry. Thymidine (blue) is functionalized with a fluorophore (red) linked at the C5 position with an azide (green)-containing cleavable linker. The ribose 3′-OH is masked by a cleavable azide (green)-containing cleavable protecting group. A Staudinger reaction using a water soluble triphenylphosphine converts the azide into hemiaminals, which upon hydrolysis releases the fluorophore and restores the free 3′-OH in a single step. This reaction product can be engaged in a new cycle of dNTP incorporation at 3′-OH.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

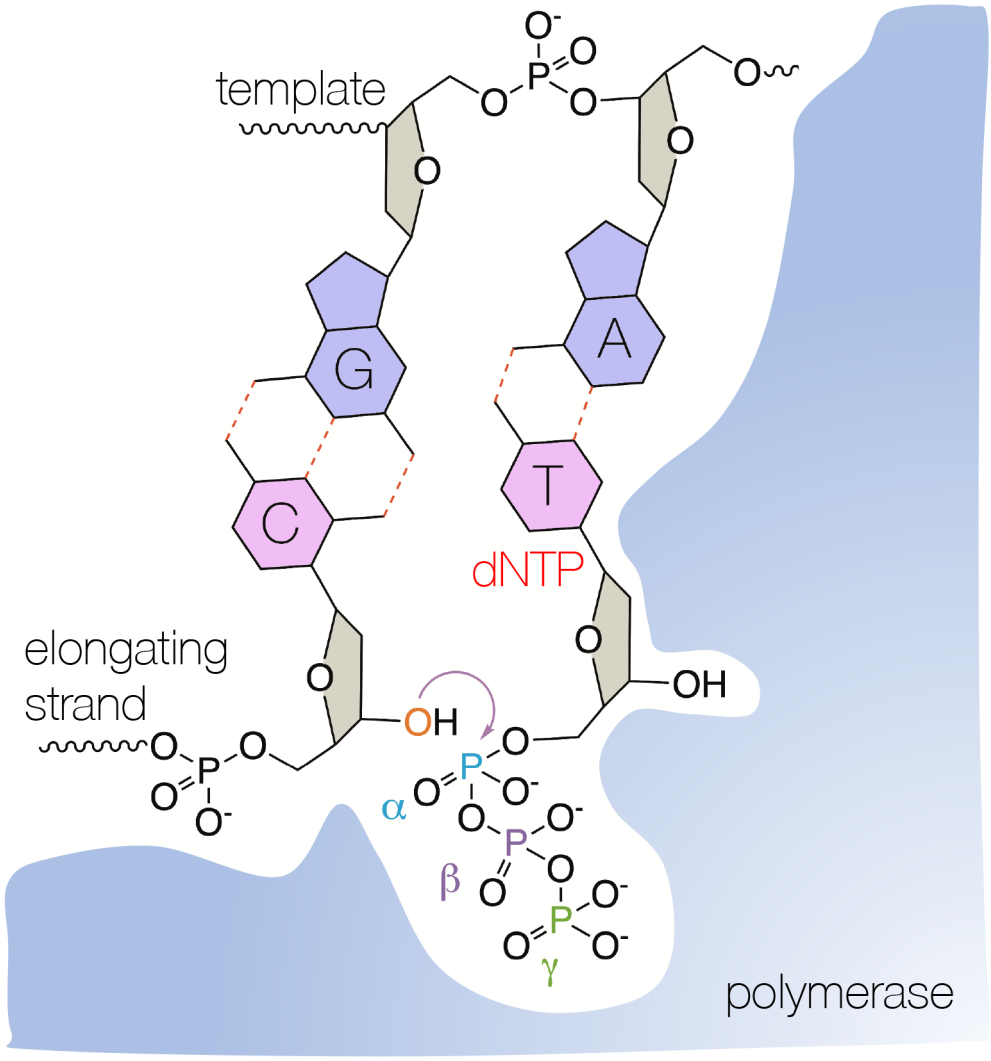

Figure 3.

Schematic illustration of the active site within a DNA polymerase showing the close positioning of the 3′-OH (brown) of the elongating strand in line with the αP (cyan) of the incoming dNTP.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

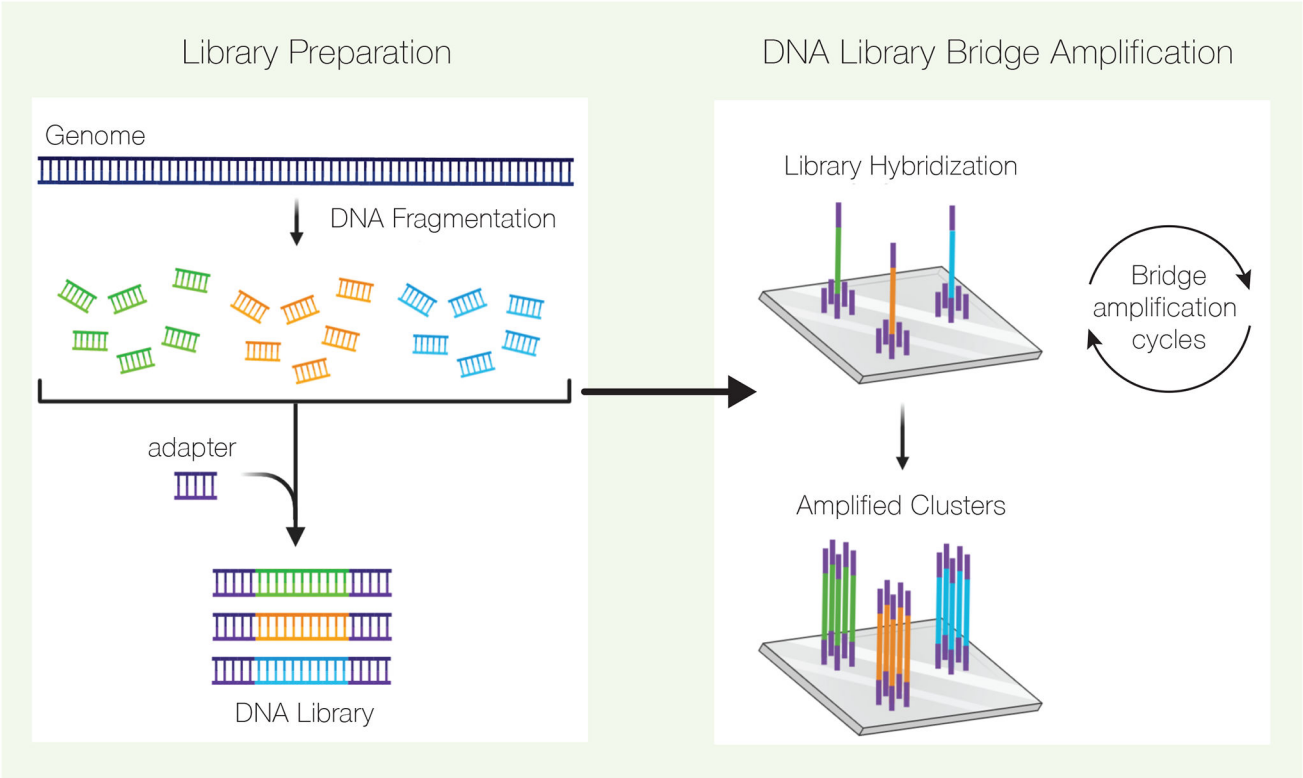

Figure 4.

Schematic of the workflow of NGS (Part 1). DNA libraries prepared by fragmenting genomic DNA into 200–300 base pairs to which adapters are attached. The library is amplified by PCR, and unique strands are hybridized on a solid support. Bridge amplification results in distinct colonies of identical strands.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

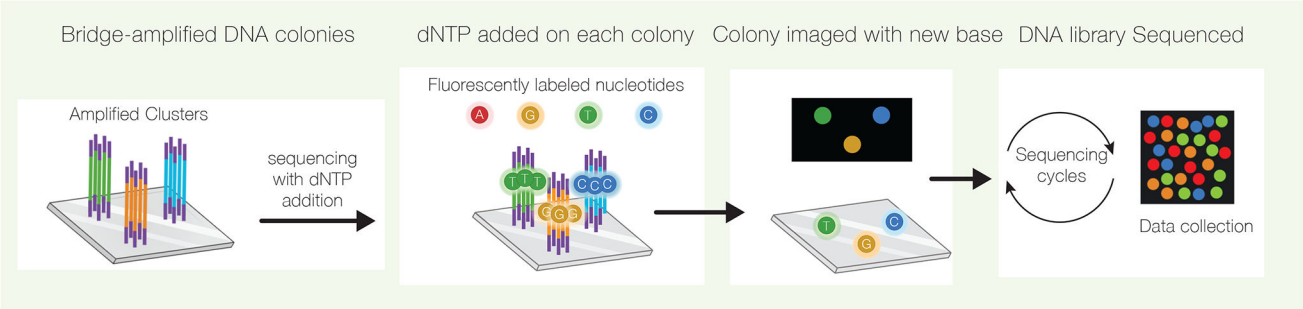

Figure 5.

Schematic of the workflow of NGS (Part 2). The distinct colonies of identical strands are then subjected to SBS and imaged in a microfluidic flow-cell using the chemistry developed by Solexa.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Technological challenges that were surmounted by new chemistries.

Challenges	Chemistry/Analysis
Detecting the incoming dNTP in the elongating DNA strand	• Design of four unnatural dNTPs each bearing a spectrally distinct, yet cleavable, fluorophore. • Integrating single molecule fluorescence microscopy and a computational data acquisition interface.
Controllable, base-by-base incorporation of dNTPs	Inventing a reversible-terminator chemistry based on a 3′-OH protecting group tolerated by DNA polymerases and cleavable under bio-orthogonal conditions.
Removal of the DNA polymerase from the DNA template after every dNTP addition for fluorescence detection and resetting the template to add the next dNTP.	Engineering a mutated DNA polymerase with a high off-rate for DNA binding.
A way to correct for the stochastic mis-incorporation of incorrect dNTPs by DNA polymerase.	Bridge-amplification chemistry to build clusters of identical DNA molecules where the major signal, from correctly incorporated dNTPs, eclipses that of a mis-incorporated dNTP.
Increase the signal-to-noise ratio of dNTP fluorescence locally after incorporation.	• Bridge-amplification chemistry on each DNA strand to create clonal, spatially separated clusters, on a solid support. • Devising a surface chemistry with low background fluorescence that could robustly anchor DNA strands to a solid support and stable over repeated cycles of dNTP addition.