

The chemistry of next-generation sequencing

Received: 3 February 2023

Accepted: 14 September 2023

Published online: 16 October 2023

 Check for updatesRaphaël Rodriguez¹✉ & Yamuna Krishnan^{2,3,4}✉

The first large genome fully sequenced by next-generation sequencing (NGS) was that of a bacteriophage using sequencing by synthesis (SBS) as a paradigm. SBS in NGS is underpinned by ‘reversible-terminator chemistry’. To grow from proof of concept to being both affordable and practical, SBS needed to overcome a series of challenges, each of which required the invention of new chemistries. These included the design and synthesis of unnatural deoxynucleotide triphosphates (dNTPs), engineering a suitable polymerase, a new surface chemistry and an ingenious molecular solution to neutralize copying errors inherent to all polymerases. In this historical Perspective, we discuss how NGS was developed from Sanger sequencing, highlighting the chemistry behind this technology, which has impacted biology in unprecedented ways.

Sanger sequencing is based on the synthesis of truncated DNA fragments from multiple DNA templates in solution. These truncations arise as a result of elongating strands prematurely terminating because of stochastic incorporation of labeled dideoxynucleotide triphosphates (ddNTPs) lacking a 3′-OH (ref. 1). This is the earliest form of ‘terminator chemistry’, in which truncated synthetic DNA strands are size separated by gel electrophoresis, after which the original DNA sequence is reconstructed (Fig. 1). Sequencing by electrophoresis and similar size-based separation of nucleic acid fragments is both time consuming and labor intensive. To sequence large genomes, a different paradigm was needed.

In 1990, Tsien and coworkers outlined a series of strategies in a patent that conceptualized SBS using automated instruments, which was abandoned shortly thereafter². They found that protection of the 3′-OH of a dNTP with a reporter group enabled stepwise incorporation by a polymerase, base by base, using an immobilized DNA template. This provided a basis for SBS technologies. Pyrosequencing, which was another early version of SBS, based on inorganic phosphate detection by luminometry, was also reported in 1998 (ref. 3).

The first map of a human genome, established using Sanger sequencing, was reported in 2001. It took 13 years, a transcontinental effort and an excess of \$0.5 billion to accomplish the first human genome map using this method^{4,5}. With NGS, a human genome can now be sequenced for \$200 within ~2 d. In this new sequencing paradigm,

short DNA reads were obtained from fragments of genomic DNA, immobilized on a solid surface and served as templates for a polymerase. Each of these fragments would be replicated (synthesized) in parallel, and the elongating strands would be 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.

NGS was developed by Solexa and is popularly known as ‘Illumina sequencing’. It gained momentum in 2007 with the first reports using chromatin immunoprecipitation coupled to sequencing^{6,7}. From 2005 to 2008, sequencing technologies distinct from NGS also materialized. These third-generation (‘post-light’) sequencing methods have emerged as complementary alternatives to NGS. Notably, Pacific Biosciences and Oxford Nanopore Technologies developed single-molecule real-time sequencing and nanopore sequencing, respectively. Conceptually different from NGS, these methods are based on electronic analysis of molecules and are true single-molecule analysis technologies. While other techniques may be more expensive, lower throughput or less accurate than NGS, they may enable sequencing long reads more effectively and are described in greater detail elsewhere^{8–10}.

The first genome sequenced using Solexa’s approach was that of the ϕ X174 bacteriophage in 2005, which demonstrated an advancement over Sanger sequencing by generating substantially more data

¹Institut Curie, CNRS, INSERM, PSL Research University, Equipe Labellisée Ligue Contre le Cancer, Paris, France. ²Department of Chemistry, the University of Chicago, Chicago, IL, USA. ³The Neuroscience Institute, the University of Chicago, Chicago, IL, USA. ⁴Institute of Biophysical Dynamics, the University of Chicago, Chicago, IL, USA. ✉e-mail: raphael.rodriguez@curie.fr; yamuna@uchicago.edu

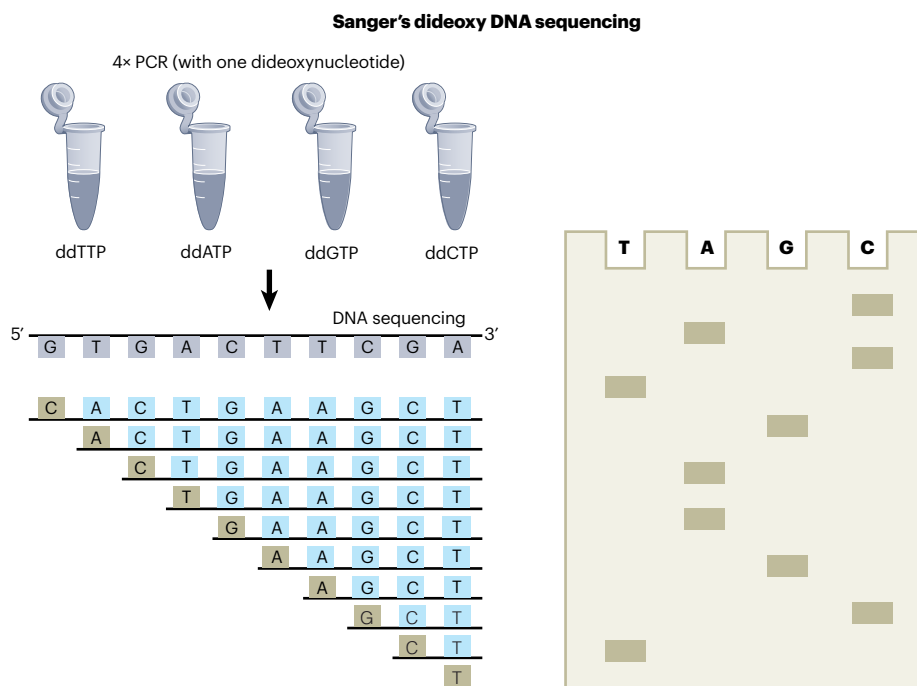


Fig. 1 | Schematic illustration of Sanger sequencing. DNA strands are separated based on size, which is defined by where into the chain a ddNTP is incorporated. This reveals the identity of the nucleobase at any position, as incorporation of a ddNTP terminates the growing chain at that position.

in a single run. Solexa launched the Genome Analyzer 1st Generation (GA1), its first commercial device. Shortly thereafter, it was acquired by Illumina, after which hundreds of scientists further improved the technology. The first human genomes were then sequenced using Solexa–Illumina NGS technology^{11,12}.

Leading up to the development of Solexa–Illumina NGS, Balasubramanian and coworkers had investigated fluorescence energy transfer between the Klenow fragment of DNA polymerase and a series of DNA primer templates, which positionally mapped nucleobases along the DNA helix at the single-molecule level¹³. The success of this experiment led to their idea that single-molecule fluorescence 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 deoxyribose to enable 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, enabling fluorescence detection and cleavage 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 rectify potential errors due to dNTP misincorporation by the polymerase.

In this historical Perspective, we detail the chemistries that were invented to establish each of the key elements of this technology and illustrate why they were instrumental to the success of SBS-based NGS as we know it today. It took several years to achieve the overall goal of successful SBS, requiring scientists with distinct knowledge spanning structural biology, biochemistry, chemical synthesis, surface chemistry, optical physics and computing sciences to explore numerous potential solutions before discovering one that worked. The current form of SBS-based DNA-sequencing

technology is the result of solving a series of technological challenges with new chemistries.

Developing an enabling reversible-terminator chemistry

One of the first challenges came to light in the 1990s, as several groups were seeking to develop new ways to make Sanger sequencing and SBS more efficient, each exploring different chemistries to this end. Many of these chemistries had their shortcomings. Early work focused on designing fluorophores at the 3′-OH of dNTPs to act as a protecting group for terminator chemistry, in which truncated synthetic DNA strands are size separated by gel electrophoresis, from which the original sequence is reconstructed, and to then use the fluorescent marker to identify the incorporated nucleotide¹⁴. While conceptually attractive, it was later found that the 3′-OH points toward amino acid residues lining the enzymatic pocket of DNA polymerases¹⁵. 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 by various types of linkers¹⁴, this requires additional rounds of cleavage, which renders it ill suited for sequencing large genomes. For example, a time of 2 h per deprotection would require 25 d to sequence a DNA fragment just 300 bases long. It was quickly realized that fluorescently labeled dNTPs at the 3′-OH were not viable.

Several groups explored the use of protecting groups at the 3′-OH for terminator chemistry¹⁶. For example, the use of 2-nitrophenol was appealing because of the fast cleavage rate upon exposure to ultraviolet light, leaving the 3′-OH available for the next round of synthesis. However, the use of ultraviolet light was not ideal, given that DNA is prone to photodamage. Repetitive cycles of irradiation can chemically alter the DNA being sequenced.

Alternatively, it was found that tethering a bulky fluorophore on the nucleobases fortuitously acted as a terminating agent by steric hindrance, obviating the need to protect the 3′-OH (ref. 17). Chemical removal of the fluorophore following its incorporation on the strand

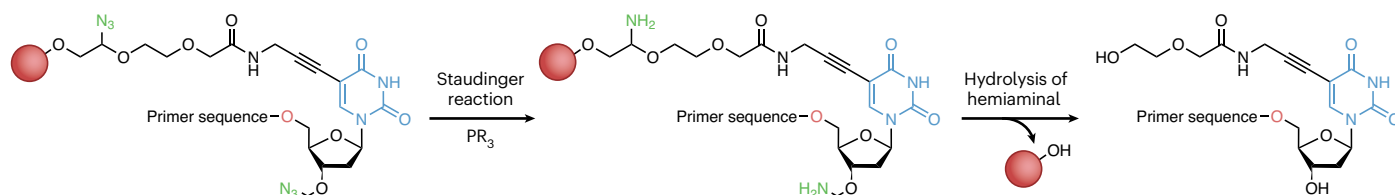


Fig. 2 | Scheme of reversible-terminator chemistry. Thymidine (blue) is functionalized with a fluorophore (red) linked at the C5 position with a cleavable linker having an azide (green). The ribose 3'-OH is masked by an azide-containing cleavable protecting group. A Staudinger reaction using a

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 the 3'-OH.

then allowed the next nucleotide to be incorporated. However, using a single-color readout, such as Alexa Fluor 594, to label all four dNTPs required the DNA template 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 increased the number of steps of the sequencing procedure and furthermore was prone to misincorporation of dNTPs due to the noncompetitive chemistry at work. If there was an absence of the correct dNTP, polymerases incorporated those available instead. This meant that the reversible-terminator chemistry and fluorescence detection had to be uncoupled from one another.

A viable strategy for addressing this limitation required a protecting group for the 3'-OH that was structurally minimalistic so that the dNTPs would remain competent substrates for DNA polymerases, highly stable under the aqueous conditions used for DNA synthesis, and cleavable in quantitative yields with a fast turnover, and would function under reaction conditions that preserved DNA integrity. An azide-based protecting group, a non-naturally occurring building block, was used to cap the 3'-OH. Its removal could be triggered by a water-soluble phosphine through the Staudinger reaction^{11,18,19}. This bio-orthogonal strategy enabled accurate base-by-base incorporation of dNTPs using differently labeled fluorescent dNTPs, bearing fluorophores on the base itself as opposed to the sugar moiety (Fig. 2). This discovery was a first step toward developing NGS as we know it today. The invention of reversible-terminator chemistry and a four-color fluorescence detection readout enabled SBS without the need to perform electrophoresis and was therefore much faster and more accurate.

Developing fluorescently labeled dNTPs for detection

Along with coworkers at Solexa, Balasubramanian and Klenerman built upon this chemistry and deduced that a working solid-surface-based NGS technology required four 'reversible-terminator' dNTPs that were designed such that each nucleobase displayed a distinct cleavable fluorophore²⁰. This would reveal the DNA code as the polymerase elongated 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^{21,22}.

The rate of polymerase elongation was a potential limitation, as it is affected by large fluorophores as well as misincorporation of unnatural dNTPs. Furthermore, it was desirable for the dNTPs to have spectrally distinct fluorescence emissions. The second major chemical challenge was to develop four different fluorophores that could be used on a solid surface. The team at Solexa successfully designed four dNTPs, each displaying a distinct fluorophore tethered by a cleavable linker that could be removed after each round of dNTP incorporation. The fluorescent guanine nucleobase, in particular, was difficult to develop, as guanine has a propensity to quench fluorophores due to single-electron transfer resulting from its low redox potential. Therefore, the fluorescent GTP monomer was designed to exhibit a different chemistry to neutralize

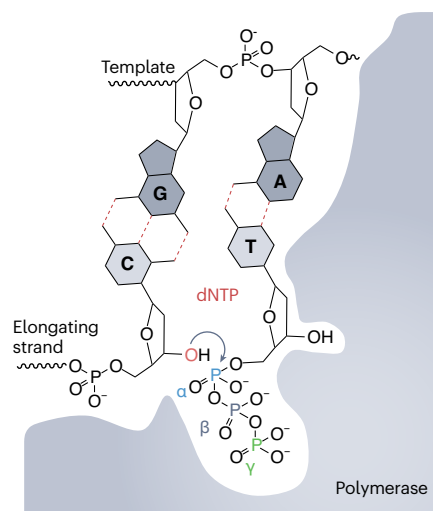


Fig. 3 | Schematic illustration of the active site within DNA polymerase.

Note the close positioning of the 3'-OH (brown) of the elongating strand in line with the α-phosphate (cyan) of the incoming dNTP.

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 DNA¹¹.

Engineering a suitable DNA polymerase for SBS

SBS operates through reversible-terminator chemistry, which enables incorporation of dNTPs base by base. Natural polymerases have evolved to incorporate natural dNTPs without pausing after each dNTP incorporation and with high speed. Getting the nucleotide chemistry correct was only one part of the challenge for DNA sequencing. Another key challenge was developing a polymerase that could accommodate this new chemistry. At that time, neither the atomic-level structure of nucleotide incorporation nor the structure of the active site was yet known, which limited our understanding of how the polymerase worked. From mechanistic insights, only a hazy map of the major residues and dynamics within the active site of the polymerase was available then.

The polymerase developed at Solexa had to tolerate the 3'-OH-protecting group better than wild-type polymerases while also exhibiting comparable kinetics of incorporation. Previous work had shown that polymerases could tolerate modifications at the 3'-OH, but nucleotide incorporation rates were slow and high concentrations of the modified nucleotides were required². Solexa's polymerase needed to be as accurate with unnatural dNTPs as the natural DNA polymerase was with natural ones. Thus, the chemistry was designed to have each of the four dNTPs compete for the enzyme active site. The accuracy

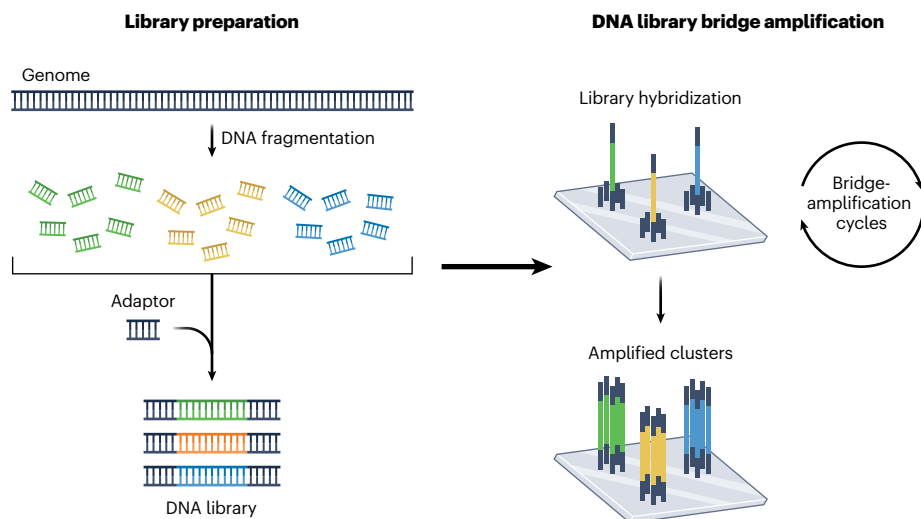


Fig. 4 | Schematic of the NGS workflow (part 1). DNA libraries are prepared by fragmenting genomic DNA into fragments of 200–300 bp to which adaptors are attached. The library is amplified by PCR, and unique strands are hybridized on a solid support. Bridge amplification results in distinct colonies of identical strands. Adapted with permission from ref. 43, Wiley.

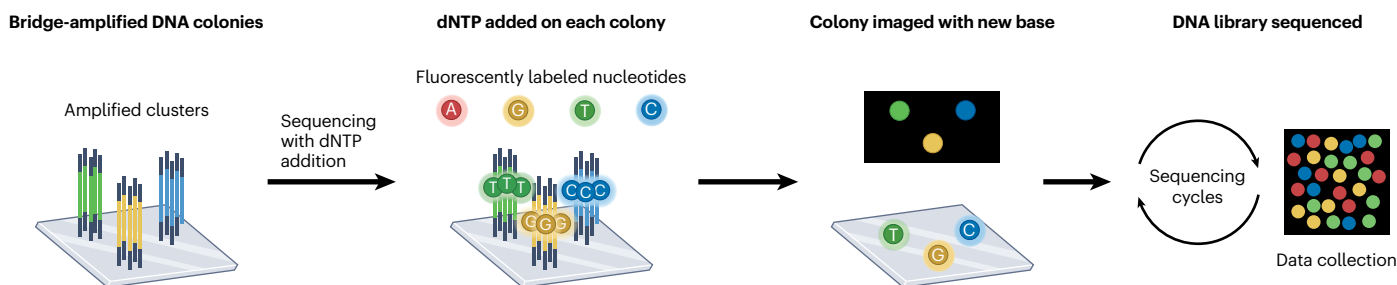


Fig. 5 | Schematic of the NGS workflow (part 2). Distinct colonies of identical strands are then subjected to SBS and imaged in a microfluidic flow cell using the chemistry developed by Solexa. Adapted with permission from ref. 43, Wiley.

stemmed from having the right dNTP in addition to the wrong ones in the sequencing reaction vessel to reduce misincorporation. There are a number of steps that guide accuracy when a polymerase incorporates a base onto a growing strand²³.

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 α -phosphate atom of the dNTP and the oxygen bridging to the β - γ -phosphates. This allows in-line nucleophilic attack of the 3'-OH that leads to an inversion at the α -phosphate (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 backward and bind a different dNTP. To optimize both steps with modified dNTPs, different polymerases were engineered and screened. Using saturation mutagenesis on three key amino acids on different polymerases, a suitable enzyme using chemically modified dNTPs was selected. This mutated polymerase could accommodate and incorporate all four unnatural dNTPs.

The polymerase also needed to fall off the DNA strand 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. Because 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.

The off-rate needed to be increased without affecting anything else on the polymerase, including the achievement of complete incorporation of the dNTP on each elongating strand before fluorescence acquisition and removal of the protecting group. Therefore, two or three positively charged amino acids on the polymerase were mutated 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^{24–26}.

Developing a surface chemistry

To begin to sequence large genomes on practical timescales, millions of smaller but distinct fragments needed to be sequenced in parallel, and then the information from these individual fragments could be pieced together correctly to reconstruct the sequence of the whole genome. This could only be achieved on a solid support, where each fragment is investigated in spatial isolation, ruling out solution-based approaches. Development of a surface chemistry that would permit this was the next challenge faced.

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. Because SBS involves repeated cycles of dNTP incorporation to sequence each of the immobilized short reads, surfaces and DNA-attachment chemistry were developed to withstand repeated exposure to the sequencing reaction conditions and avoid the loss of DNA strands that were being sequenced. The developed surface also had to be passive to prevent background

Table 1 | Technological challenges that were surmounted by new chemistries

| Challenges | Chemistry/analysis |
|--|--|
| Detecting the incoming dNTP in the elongating DNA strand | <ul style="list-style-type: none"> • 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 | <ul style="list-style-type: none"> • 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 | <ul style="list-style-type: none"> • Engineering a mutated DNA polymerase with a high off-rate for DNA binding |
| A way to correct for stochastic misincorporation of incorrect dNTPs by DNA polymerase | <ul style="list-style-type: none"> • Bridge-amplification chemistry to build clusters of identical DNA molecules where the major signal from correctly incorporated dNTPs eclipses that of a misincorporated dNTP |
| Increase the signal-to-noise ratio of dNTP fluorescence locally after incorporation | <ul style="list-style-type: none"> • 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 onto which DNA could robustly be anchored and remain stable over repeated cycles of dNTP addition |

fluorescence that could interfere with the fluorescence of incorporated dNTPs. Another key challenge was to prevent nonspecific binding of fluorescent dNTPs to the surface, because 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 nonspecific background needed to be drastically reduced by surface engineering 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 and then amino–DNA grafting. Polyacrylamide surface chemistry offered more stability to anchored DNA strands and a higher signal-to-noise ratio²⁷. Illumina sequencing platforms have evolved tremendously over the past 2 decades, and the surface chemistry used in the latest apparatus has, to our knowledge, not been reported.

Enabling parallelization for speed and accuracy

Detection of fluorophores with extreme sensitivity was another obvious challenge. At this point, it was clear that sequencing unique copies of DNA fragments would invariably yield a low signal-to-noise ratio. Solexa therefore needed to deviate from their strategy as initially conceived and incorporated a technology developed at Manteia Predictive Medicine by Mayer and coworkers²⁸. 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. Going from single-molecule fluorescence to 'few-molecule fluorescence' had two major advantages. It improved the signal-to-noise ratio substantially. 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.

In molecular clustering, specific primers are attached on both extremities of every genomic DNA fragment subjected to sequencing. These modified DNA fragments are then hybridized to the surface, which displays DNA sequences complementary to these primers. Amplification with DNA polymerase results in 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, also known as 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

implementing an isothermal amplification on the surface²⁹. 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 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 the flow cell onto which the DNA to be sequenced is loaded was mobile, so that large areas could be imaged and the sequencing could be massively parallelized (Fig. 5)^{11,12}. Multiplying fragments in this manner enabled acquisition of signal a thousand times faster.

Conclusions

For over a century, chemists have attempted to build complex, naturally occurring molecules by mimicking reactions and using substrates found in nature^{30,31}, with the idea that biomimetic strategies are elegant and straightforward and have greater 'atom economy' than those that rely on protecting groups^{32–34}. Indeed, DNA polymerases synthesize DNA without protecting groups. In the context of NGS, however, it is the introduction of a protecting group and a chemistry not found in nature that enables 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.

Here we have described the key elements that went into chemically developing a technology that turned a pipe dream into reality. This included the rational design and challenging synthesis of fluorescently labeled and protected unnatural dNTPs, a new DNA polymerase, a new surface chemistry and a method to locally amplify unique copies of the DNA fragment to be sequenced to maximize sequencing accuracy (Table 1). NGS has been used across the spectrum of life sciences because the ease of DNA sequencing provides the means to interrogate biology and to solve medical problems in unparalleled ways. It has enabled the identification of risk genes for diseases such as autism for which the underlying genetics was unclear³⁵. It illuminated chromatin biology, providing insights into the genomic locations of chromatin marks⁶, enabling genome-wide mapping of protein–DNA interactions⁷ and establishing genomic sites of action of small molecules^{36,37}. It has been used as a molecular counter to quantify RNA transcripts in tissues with spatial resolution^{38,39}. NGS impacted clinical medicine by revealing disease-causing mutations³⁵, providing critical information that guided therapeutic strategies and enabling non-invasive circulating cell-free DNA analysis as diagnostic markers⁴⁰. More recently, NGS made in vivo genome editing possible by providing sequence maps of genomic loci that can be altered and to validate selective editing⁴¹. In addition, ancestry can be established from analyses of big data, cold

cases have been solved and innocents on death row have been exonerated⁴⁰. It was used to decipher viral genomes, which was instrumental in the rapid development of RNA vaccines⁴². Specifically applied to severe acute respiratory syndrome coronavirus 2 to track variants of concern, along with the RNA vaccines, NGS formed the bedrock of our response to the pandemic. Finally, NGS is reshaping our understanding of genome evolution, providing insights into future evolutionary trends across species.

From the breadth of its impact on biology, medicine and public health, NGS illustrates how central chemistry is to the invention of new technologies. The development of NGS is the result not only of thinking outside the box, risk taking and many failed experiments but of the realization up front of what would make it a successful technology and the capacity to identify key problems to be solved along the way. Since its inception, NGS has kept improving in terms of increasing sequencing speed and lowering costs; however, no new chemistries have been disclosed. Some of the major early improvements were made on DNA library preparation. Other technologies have emerged since, providing powerful alternatives in particular for the sequencing of longer DNA fragments. Future efforts may be focused on data analysis in this new era in which we scientists and clinicians can produce more data than can be analyzed.

References

- Sanger, F., Nicklen, F. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
- Tsien, R. Y., Ross, P., Fahnestock, M. & Johnston, A. DNA sequencing. Patent WO9106678A1 (1990).
- Ronaghi, M., Uhlén, M. & Nyrén, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
- International Human Genome Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Johnson, G. D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
- Gupta, N. & Verma, V. K. Next-generation sequencing and its application: empowering in public health beyond reality. In *Microbial Technology for the Welfare of Society: Microorganisms for Sustainability* (ed. Arora, P.) Vol. 17 (Springer, 2019).
- Willson, J. Sequencing — the next generation. *Nature Milestones* S7 (2021).
- Burgess, D. The dawn of personal genomes. *Nature Milestones* S9 (2021).
- Bentley, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Wang, J. et al. The diploid genome sequence of an Asian individual. *Nature* **456**, 60–66 (2008).
- Furey, W. S. et al. Use of fluorescence resonance energy transfer to investigate the conformation of DNA substrates bound to the Klenow fragment. *Biochemistry* **37**, 2979–2990 (1998).
- Canard, B. & Sarfati, R. S. DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene* **148**, 1–6 (1994).
- Welch, M. B. et al. Syntheses of nucleosides designed for combinatorial DNA sequencing. *Chem. Eur. J.* **5**, 951–960 (1999).
- Metzker, M. L. et al. Termination of DNA synthesis by novel 3'-modified deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.* **22**, 4259–4267 (1994).
- Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A.-P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* **36**, e25 (2008).
- Staudinger, H. & Meyer, J. Über neue organische phosphorverbindungen III. Phosphinmethylesterivate und phosphinimine. *Helv. Chim. Acta* **2**, 635 (1919).
- Balasubramanian, S. Sequencing nucleic acids: from chemistry to medicine. *Chem. Commun.* **47**, 7281–7286 (2011).
- Barnes, C., Balasubramanian, S., Liu, X., Swerdlow, H. & Milton, J. Labelled nucleotides. US patent 7,057,026 B2 (2002).
- Sarfati, S. R. et al. Synthesis of fluorescent or biotinylated nucleoside compounds. *Tetrahedron* **43**, 3491–3497 (1987).
- Rosenblum, B. B. et al. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**, 4500–4504 (1997).
- Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307 (1998).
- Smith, G. P., Bailey, D. M. D., Sanches, R. M., Swerdlow, H. & Earnshaw, D. J. Modified polymerases for improved incorporation of nucleotide analogues. Patent WO 2005/024010 A1 (2003).
- Ost, T. W. B., Smith, G. P., Balasubramanian, S., Rigatti, R. & Sanches, R. M. Improved polymerases. Patent WO 2006/120433 A1 (2006).
- Smith, G. P., Bailey, D. M. D., Sanches-Kuiper, R. M., Swerdlow, H. & Earnshaw, D. J. Modified polymerases for improved incorporation of nucleotide analogues. US patent 8,852,910 B2 (2003).
- Smith, M. et al. Modified molecular arrays. Patent WO 2005/065814 A1 (2005).
- Kawashima, E., Farinelli, L. & Mayer, P. Method of nucleic acid amplification by extension of immobilized primers. Patent WO 9844151 (1998).
- Mayer, P. Isothermal amplification of nucleic acids on a solid support. Patent WO 02/46456 (2001).
- Robinson, R. A synthesis of tropinone. *J. Chem. Soc.* **111**, 762–768 (1917).
- Rodriguez, A. R. et al. Total synthesis of cyercene A and the biomimetic synthesis of (±)-9,10-deoxytridachione and (±)-ocellapyrone. *Tetrahedron* **63**, 4500–4509 (2007).
- Wuts, P. G. M. & Greene, T. W. *Greene's Protective Groups in Organic Synthesis* 4th edn (John Wiley & Sons, 2006).
- Baran, P. S., Maimone, T. J. & Richter, J. M. Total synthesis of marine natural products without using protecting groups. *Nature* **446**, 404–408 (2007).
- Trost, B. M. The atom economy—a search for synthetic efficiency. *Science* **254**, 1471–1477 (1991).
- Koboldt, D. C., Meltz Steinberg, K., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38 (2013).
- Rodriguez, R. et al. Small molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
- Rodriguez, R. & Miller, K. M. Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nat. Rev. Genet.* **15**, 783–796 (2014).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).
- Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Shendure, J. et al. DNA sequencing at 40: past, present and future. *Nature* **550**, 346–353 (2017).
- Newby, G. A. et al. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295–302 (2021).

42. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
43. Veetil, A. T. & Krishnan, Y. In *Advanced Chemical Biology* (eds Hang, H. C., Pratt, M. R. & Prescher, J. A.) Ch. 2, pp 9–30 (Wiley-VCH, 2023).

Acknowledgements

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.

Author contributions

R.R. and Y.K. contributed equally to this work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Raphaël Rodriguez or Yamuna Krishnan.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2023