Exhibit L

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Sequencing technologies — the next generation

*Michael L. Metzker**‡

Abstract | Demand has never been greater for revolutionary technologies that deliver fast, inexpensive and accurate genome information. This challenge has catalysed the development of next-generation sequencing (NGS) technologies. The inexpensive production of large volumes of sequence data is the primary advantage over conventional methods. Here, I present a technical review of template preparation, sequencing and imaging, genome alignment and assembly approaches, and recent advances in current and near-term commercially available NGS instruments. I also outline the broad range of applications for NGS technologies, in addition to providing guidelines for platform selection to address biological questions of interest.

Automated Sanger sequencing
This process involves a mixture of techniques: bacterial cloning or PCR; template purification; labelling of DNA fragments using the chain termination method with energy transfer, dye-labelled dideoxynucleotides and a DNA polymerase; capillary electrophoresis; and fluorescence detection that provides four-colour plots to reveal the DNA sequence.

**Human Genome Sequencing Center and Department of Molecular & Human Genetics, Baylor College of Medicine, One Baylor Plaza, N1409, Houston, Texas 77030, USA.*
‡LaserGen, Inc., 8052 El Rio Street, Houston, Texas 77054, USA.
e-mail: mmetzker@bcm.edu
doi:10.1038/nrg2626
Published online
8 December 2009

Over the past four years, there has been a fundamental shift away from the application of automated Sanger sequencing for genome analysis. Prior to this departure, the automated Sanger method had dominated the industry for almost two decades and led to a number of monumental accomplishments, including the completion of the only finished-grade human genome sequence[1]. Despite many technical improvements during this era, the limitations of automated Sanger sequencing showed a need for new and improved technologies for sequencing large numbers of human genomes. Recent efforts have been directed towards the development of new methods, leaving Sanger sequencing with fewer reported advances. As such, automated Sanger sequencing is not covered here, and interested readers are directed to previous articles[2,3].

The automated Sanger method is considered as a 'first-generation' technology, and newer methods are referred to as next-generation sequencing (NGS). These newer technologies constitute various strategies that rely on a combination of template preparation, sequencing and imaging, and genome alignment and assembly methods. The arrival of NGS technologies in the marketplace has changed the way we think about scientific approaches in basic, applied and clinical research. In some respects, the potential of NGS is akin to the early days of PCR, with one's imagination being the primary limitation to its use. The major advance offered by NGS is the ability to produce an enormous volume of data cheaply — in some cases in excess of one billion short reads per instrument run. This feature expands the realm of experimentation beyond just determining the order of bases. For example, in gene-expression studies microarrays are now being replaced by seq-based methods, which can identify and quantify rare transcripts without prior knowledge of a particular gene and can provide information regarding alternative splicing and sequence variation in identified genes[4,5]. The ability to sequence the whole genome of many related organisms has allowed large-scale comparative and evolutionary studies to be performed that were unimaginable just a few years ago. The broadest application of NGS may be the resequencing of human genomes to enhance our understanding of how genetic differences affect health and disease. The variety of NGS features makes it likely that multiple platforms will coexist in the marketplace, with some having clear advantages for particular applications over others.

This Review focuses on commercially available technologies from Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences, the Polonator instrument and the near-term technology of Pacific Biosciences, who aim to bring their sequencing device to the market in 2010. Nanopore sequencing is not covered, although interested readers are directed to an article by Branton and colleagues[6], who describe the advances and remaining challenges for this technology. Here, I present a technical review of template preparation, sequencing and imaging, genome alignment and assembly, and current NGS platform performance to provide guidance on how these technologies work and how they may be applied to important biological questions. I highlight the applications of human genome resequencing using targeted and whole-genome approaches, and discuss the progress

©2010 Macmillan Publishers Limited. All rights reserved

and limitations of these methods, as well as upcoming advances and the impact they are expected to have over the next few years.

Next-generation sequencing technologies

Sequencing technologies include a number of methods that are grouped broadly as template preparation, sequencing and imaging, and data analysis. The unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform. These differences in data output present challenges when comparing platforms based on data quality and cost. Although quality scores and accuracy estimates are provided by each manufacturer, there is no consensus that a 'quality base' from one platform is equivalent to that from another platform. Various sequencing metrics are discussed later in the article.

In the following sections, stages of template preparation and sequencing and imaging are discussed as they apply to existing and near-term commercial platforms. There are two methods used in preparing templates for NGS reactions: clonally amplified templates originating from single DNA molecules, and single DNA-molecule templates. The term sequencing by synthesis, which is used to describe numerous DNA polymerase-dependent methods in the literature, is not used in this article because it fails to delineate the different mechanisms involved in sequencing[2,7]. Instead, these methods are classified as cyclic reversible termination (CRT), single-nucleotide addition (SNA) and real-time sequencing. Sequencing by ligation (SBL), an approach in which DNA polymerase is replaced by DNA ligase, is also described. Imaging methods coupled with these sequencing strategies range from measuring bioluminescent signals to four-colour imaging of single molecular events. The voluminous data produced by these NGS platforms place substantial demands on information technology in terms of data storage, tracking and quality control (see REF. 8 for details).

Template preparation

The need for robust methods that produce a representative, non-biased source of nucleic acid material from the genome under investigation cannot be overemphasized. Current methods generally involve randomly breaking genomic DNA into smaller sizes from which either fragment templates or mate-pair templates are created. A common theme among NGS technologies is that the template is attached or immobilized to a solid surface or support. The immobilization of spatially separated template sites allows thousands to billions of sequencing reactions to be performed simultaneously.

Clonally amplified templates. Most imaging systems have not been designed to detect single fluorescent events, so amplified templates are required. The two most common methods are emulsion PCR (emPCR)[9] and solid-phase amplification[10]. EmPCR is used to prepare sequencing templates in a cell-free system, which has the advantage of avoiding the arbitrary loss of genomic sequences — a problem that is inherent in bacterial cloning methods. A library of fragment or mate-pair targets is created, and adaptors containing universal priming sites are ligated to the target ends, allowing complex genomes to be amplified with common PCR primers. After ligation, the DNA is separated into single strands and captured onto beads under conditions that favour one DNA molecule per bead (FIG. 1a). After the successful amplification and enrichment of emPCR beads, millions can be immobilized in a polyacrylamide gel on a standard microscope slide (Polonator)[11], chemically crosslinked to an amino-coated glass surface (Life/APG; Polonator)[12] or deposited into individual PicoTiterPlate (PTP) wells (Roche/454)[13] in which the NGS chemistry can be performed.

Solid-phase amplification can also be used to produce randomly distributed, clonally amplified clusters from fragment or mate-pair templates on a glass slide (FIG. 1b). High-density forward and reverse primers are covalently attached to the slide, and the ratio of the primers to the template on the support defines the surface density of the amplified clusters. Solid-phase amplification can produce 100–200 million spatially separated template clusters (Illumina/Solexa), providing free ends to which a universal sequencing primer can be hybridized to initiate the NGS reaction.

Single-molecule templates. Although clonally amplified methods offer certain advantages over bacterial cloning, some of the protocols are cumbersome to implement and require a large amount of genomic DNA material (3–20 μg). The preparation of single-molecule templates is more straightforward and requires less starting material (<1 μg). More importantly, these methods do not require PCR, which creates mutations in clonally amplified templates that masquerade as sequence variants. AT-rich and GC-rich target sequences may also show amplification bias in product yield, which results in their underrepresentation in genome alignments and assemblies. Quantitative applications, such as RNA–seq[5], perform more effectively with non-amplified template sources, which do not alter the representational abundance of mRNA molecules.

Before the NGS reaction is carried out, single-molecule templates are usually immobilized on solid supports using one of at least three different approaches. In the first approach, spatially distributed individual primer molecules are covalently attached to the solid support[14]. The template, which is prepared by randomly fragmenting the starting material into small sizes (for example, ~200–250 bp) and adding common adaptors to the fragment ends, is then hybridized to the immobilized primer (FIG. 1c). In the second approach, spatially distributed single-molecule templates are covalently attached to the solid support[14] by priming and extending single-stranded, single-molecule templates from immobilized primers (FIG. 1c). A common primer is then hybridized to the template (FIG. 1d). In either approach, DNA polymerase can bind to the immobilized primed template configuration to initiate the NGS reaction. Both of the above approaches are used by Helicos BioSciences. In a third approach, spatially distributed single polymerase molecules

Finished grade
A quality measure for a sequenced genome. A finished-grade genome, commonly referred to as a 'finished genome', is of higher quality than a draft-grade genome, with more base coverage and fewer errors and gaps (for example, the human genome reference contains 2.85 Gb, covers 99% of the genome with 341 gaps, and has an error rate of 1 in every 100,000 bp).

Template
This recombinant DNA molecule is made up of a known region, usually a vector or adaptor sequence to which a universal primer can bind, and the target sequence, which is typically an unknown portion to be sequenced.

Seq-based methods
Assays that use next-generation sequencing technologies. They include methods for determining the sequence content and abundance of mRNAs, non-coding RNAs and small RNAs (collectively called RNA–seq) and methods for measuring genome-wide profiles of immunoprecipitated DNA–protein complexes (ChIP–seq), methylation sites (methyl–seq) and DNase I hypersensitivity sites (DNase–seq).

Polonator
This Review mostly describes technology platforms that are associated with a respective company, but the Polonator G.007 instrument, which is manufactured and distributed by Danaher Motions (a Dover Company), is an open source platform with freely available software and protocols. Users manufacture their own reagents based on published reports or by collaborating with George Church and colleagues or other technology developers.

Fragment templates
A fragment library is prepared by randomly shearing genomic DNA into small sizes of <1 kb, and requires less DNA than would be needed for a mate-pair library.

©2010 Macmillan Publishers Limited. All rights reserved

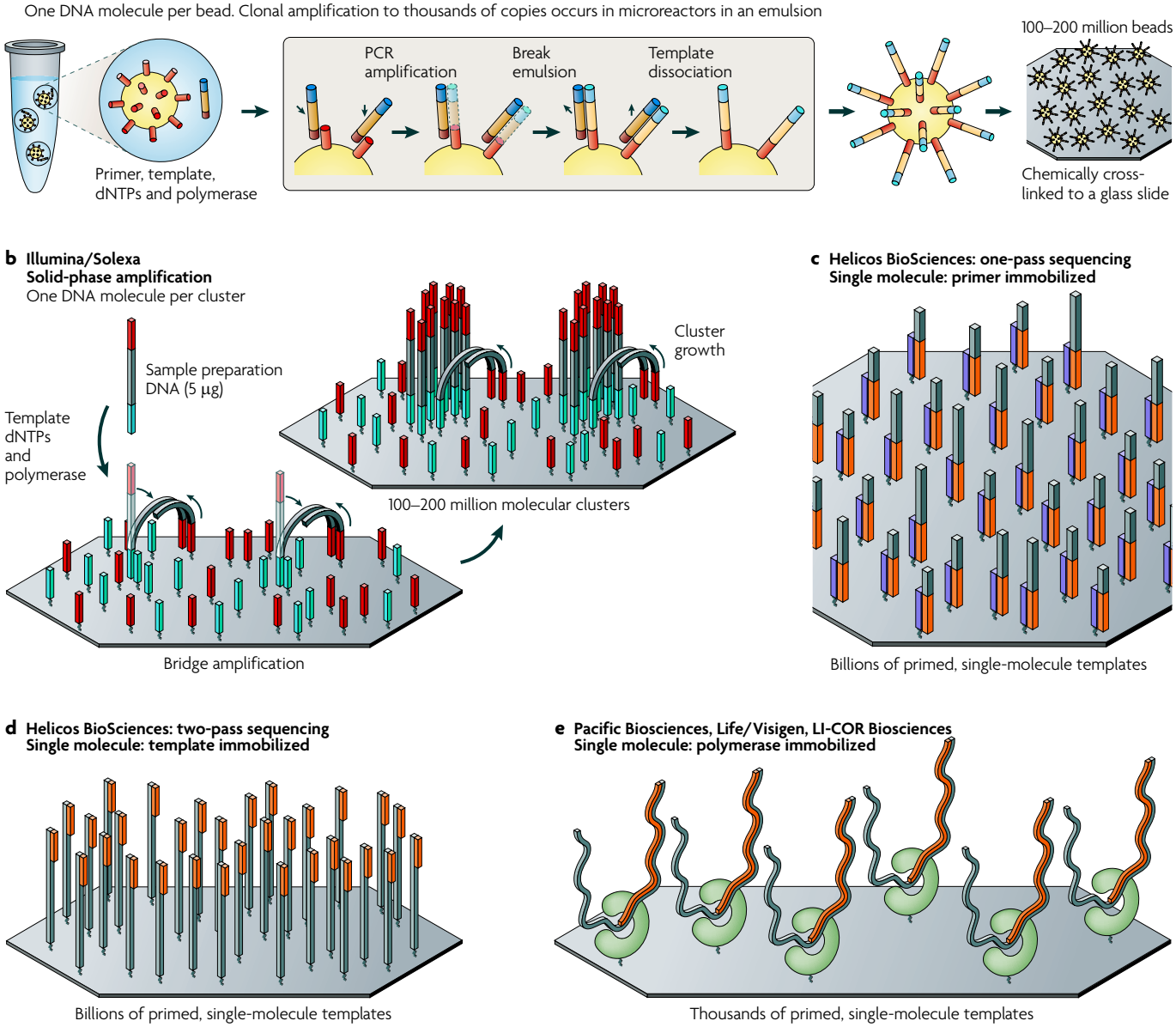


Figure 1 | **Template immobilization strategies.** In emulsion PCR (emPCR) (**a**), a reaction mixture consisting of an oil–aqueous emulsion is created to encapsulate bead–DNA complexes into single aqueous droplets. PCR amplification is performed within these droplets to create beads containing several thousand copies of the same template sequence. EmPCR beads can be chemically attached to a glass slide or deposited into PicoTiterPlate wells (FIG. 3c). Solid-phase amplification (**b**) is composed of two basic steps: initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. Three approaches are shown for immobilizing single-molecule templates to a solid support: immobilization by a primer (**c**); immobilization by a template (**d**); and immobilization of a polymerase (**e**). dNTP, 2′-deoxyribonucleoside triphosphate.

Mate-pair templates
A genomic library is prepared by circularizing sheared DNA that has been selected for a given size, such as 2 kb, therefore bringing the ends that were previously distant from one another into close proximity. Cutting these circles into linear DNA fragments creates mate-pair templates.

are attached to the solid support[15], to which a primed template molecule is bound (FIG. 1e). This approach is used by Pacific Biosciences[15] and is described in patents from Life/VisiGen[16] and LI-COR Biosciences[17]. Larger DNA molecules (up to tens of thousands of base pairs) can be used with this technique and, unlike the first two approaches, the third approach can be used with real-time methods, resulting in potentially longer read lengths.

Sequencing and imaging

There are fundamental differences in sequencing clonally amplified and single-molecule templates. Clonal amplification results in a population of identical templates, each of which has undergone the sequencing reaction. Upon imaging, the observed signal is a consensus of the nucleotides or probes added to the identical templates for a given cycle. This places a greater

©2010 Macmillan Publishers Limited. All rights reserved

Case 1:10-cv-00735-SLR Document 9-13 Filed 10/22/10 Page 5 of 8

demand on the efficiency of the addition process, and incomplete extension of the template ensemble results in lagging-strand dephasing. The addition of multiple nucleotides or probes can also occur in a given cycle, resulting in leading-strand dephasing. Signal dephasing increases fluorescence noise, causing base-calling errors and shorter reads[18]. Because dephasing is not an issue with single-molecule templates, the requirement for cycle efficiency is relaxed. Single molecules, however, are susceptible to multiple nucleotide or probe additions in any given cycle. Here, deletion errors will occur owing to quenching effects between adjacent dye molecules or no signal will be detected because of the incorporation of dark nucleotides or probes. In the following sections, sequencing and imaging strategies that use both clonally amplified and single-molecule templates are discussed.

Cyclic reversible termination. As the name implies, CRT uses reversible terminators in a cyclic method that comprises nucleotide incorporation, fluorescence imaging and cleavage[2]. In the first step, a DNA polymerase, bound to the primed template, adds or incorporates just one fluorescently modified nucleotide (BOX 1), which represents the complement of the template base. The termination of DNA synthesis after the addition of a single nucleotide is an important feature of CRT. Following incorporation, the remaining unincorporated nucleotides are washed away. Imaging is then performed to determine the identity of the incorporated nucleotide. This is followed by a cleavage step, which removes the terminating/inhibiting group and the fluorescent dye. Additional washing is performed before starting the next incorporation step. FIG. 2a depicts a four-colour CRT cycle used by Illumina/Solexa, and FIG. 2c illustrates a one-colour CRT cycle used by Helicos BioSciences.

The key to the CRT method is the reversible terminator, of which there are two types: 3′ blocked and 3′ unblocked (BOX 1). The use of a dideoxynucleotide, which acts as a chain terminator in Sanger sequencing, provided the basis for the initial development of reversible blocking groups attached to the 3′ end of nucleotides[19,20]. Blocking groups, such as 3′-*O*-allyl-2′-deoxyribonucleoside triphosphates (dNTPs)[21] and 3′-*O*-azidomethyl-dNTPs[22], have been successfully used in CRT. 3′-blocked terminators require the cleavage of two chemical bonds to remove the fluorophore from the nucleobase and restore the 3′-OH group.

Currently, the Illumina/Solexa Genome Analyzer (GA)[23] dominates the NGS market. It uses the clonally amplified template method illustrated in FIG. 1b, coupled with the four-colour CRT method illustrated in FIG. 2a. The four colours are detected by total internal reflection fluorescence (TIRF) imaging using two lasers, the output of which is depicted in FIG. 2b. The slide is partitioned into eight channels, which allows independent samples to be run simultaneously. TABLE 1 shows the current sequencing statistics of the Illumina/Solexa GA_{II} platform operating at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC; D. Muzny, personal communication). Substitutions are the most common error type, with a higher portion of errors occurring when the previous incorporated nucleotide is a 'G' base[24]. Genome analysis of Illumina/Solexa data has revealed an underrepresentation of AT-rich[24–26] and GC-rich regions[25,26], which is probably due to amplification bias during template preparation[25]. Sequence variants are called by aligning reads to a reference genome using bioinformatics tools such as MAQ[27] or ELAND[23]. Bentley and colleagues reported high concordance (>99.5%) of single-nucleotide variant (SNV)[28] calls with standard genotyping arrays using both alignment tools, and a false-positive rate of 2.5% with novel SNVs[23]. Other reports have described a higher false-positive rate associated with novel SNV detection using these alignment tools[29,30].

The difficulty involved in identifying a modified enzyme that efficiently incorporates 3′-blocked terminators — a process that entails screening large libraries of mutant DNA polymerases — has spurred the development of 3′-unblocked reversible terminators. LaserGen, Inc. was the first group to show that a small terminating group attached to the base of a 3′-unblocked nucleotide can act as an effective reversible terminator and be efficiently incorporated by wild-type DNA polymerases[31]. This led to the development of Lightning Terminators[32] (BOX 1). Helicos BioSciences has reported the development of Virtual Terminators, which are 3′-unblocked terminators with a second nucleoside analogue that acts as an inhibitor[33]. The challenge for 3′-unblocked terminators is creating the appropriate modifications to the terminating (Lightning Terminators)[32] or inhibiting (Virtual Terminators)[33] groups so that DNA synthesis is terminated after a single base addition. This is important because an unblocked 3′-OH group is the natural substrate for incorporating the next incoming nucleotide. Cleavage of only a single bond is required to remove both the terminating or inhibiting group and the fluorophore group from the nucleobase, which is a more efficient strategy than 3′-blocked terminators for restoring the nucleotide for the next CRT cycle.

Helicos BioSciences was the first group to commercialize a single-molecule sequencer, the HeliScope, which was based on the work of Quake and colleagues[34]. The HeliScope uses the single-molecule template methods shown in FIG. 1c and FIG. 1d coupled with the one-colour (Cy5 dye) CRT method shown in FIG. 2c. Incorporation of a nucleotide results in a fluorescent signal. The HeliScope also uses TIRF to image the Cy5 dye[34], the imaging output of which is shown in FIG. 2d. Harris and colleagues[14] used Cy5-12ss-dNTPs, which are earlier versions of their Virtual Terminators that lack the inhibiting group, and reported that deletion errors in homopolymeric repeat regions were the most common error type (~5% frequency) when using the primer-immobilized strategy shown in FIG. 1c. This is likely to be related to the incorporation of two or more Cy5-12ss-dNTPs in a given cycle. These errors can be greatly reduced with two-pass sequencing, which provides ~25-base consensus reads using the template-immobilized strategy shown in FIG. 1d. At the 2009 Advances in Genome Biology and Technology (AGBT) meeting, the Helicos group reported their recent progress in sequencing the

Dephasing
This occurs with step-wise addition methods when growing primers move out of synchronicity for any given cycle. Lagging strands (for example, $n - 1$ from the expected cycle) result from incomplete extension, and leading strands (for example, $n + 1$) result from the addition of multiple nucleotides or probes in a population of identical templates.

Dark nucleotides or probes
A nucleotide or probe that does not contain a fluorescent label. It can be generated from its cleavage and carry-over from the previous cycle or be hydrolysed *in situ* from its dye-labelled counterpart in the current cycle.

Total internal reflection fluorescence
A total internal reflection fluorescence imaging device produces an evanescent wave — that is, a near-field stationary excitation wave with an intensity that decreases exponentially away from the surface. This wave propagates across a boundary surface, such as a glass slide, resulting in the excitation of fluorescent molecules near (<200 nm) or at the surface and the subsequent collection of their emission signals by a detector.

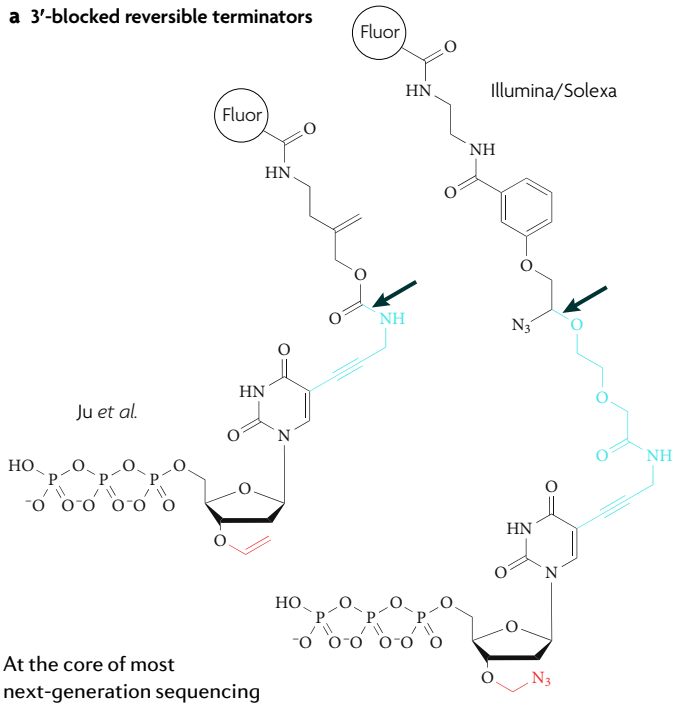
Libraries of mutant DNA polymerases
Large numbers of genetically engineered DNA polymerases can be created by either site-directed or random mutagenesis, which leads to one or more amino acid substitutions, insertions and/or deletions in the polymerase. The goal of this approach is to incorporate modified nucleotides more efficiently during the sequencing reaction.

Consensus reads
These are only useful for single-molecule techniques and are produced by sequencing the same template molecule more than once. The data are then aligned to produce a 'consensus read', reducing stochastic errors that may occur in a given sequence read.

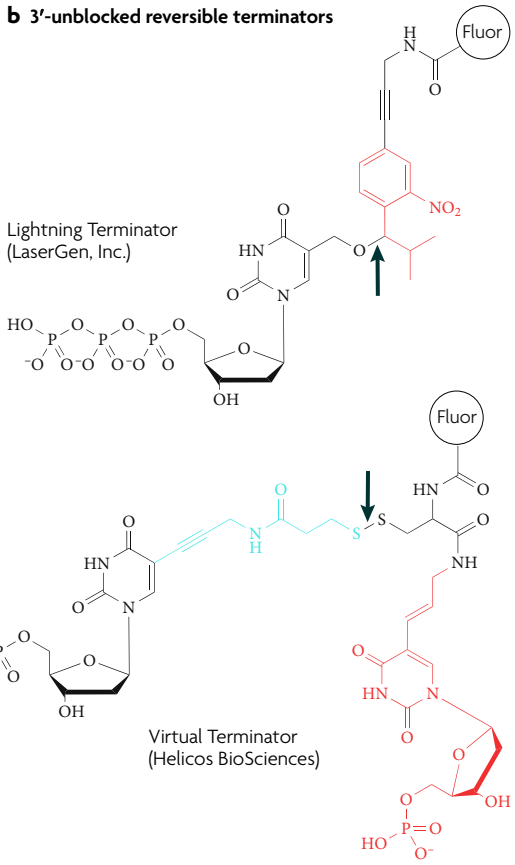
©2010 Macmillan Publishers Limited. All rights reserved

Box 1 | **Modified nucleotides used in next-generation sequencing methods**

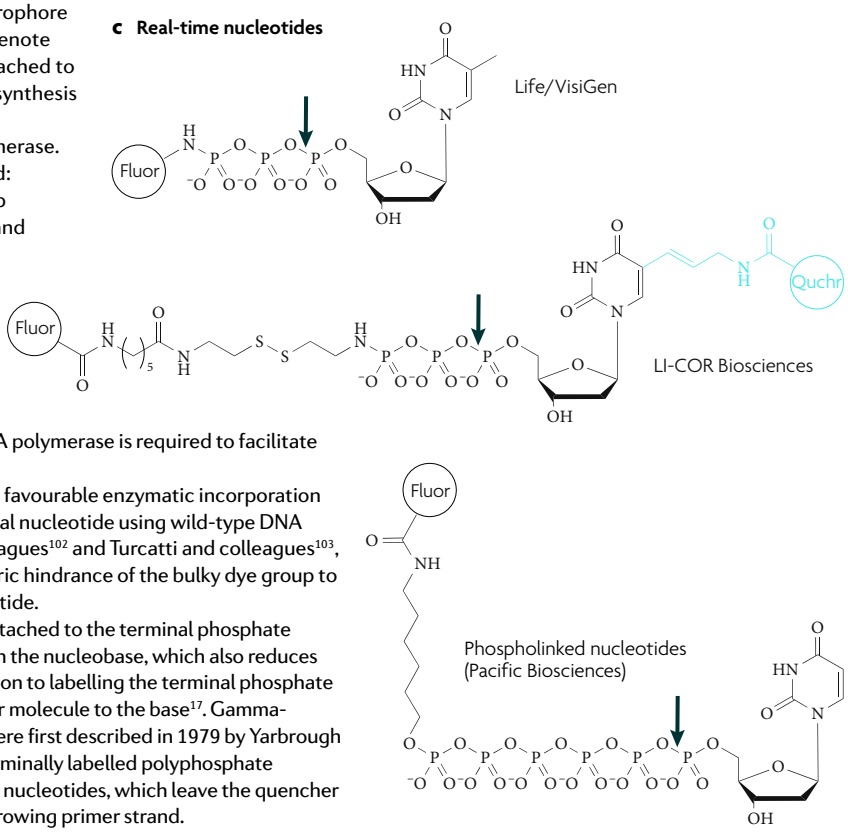
a 3′-blocked reversible terminators



b 3′-unblocked reversible terminators



c Real-time nucleotides



At the core of most next-generation sequencing (NGS) methods is the use of dye-labelled modified nucleotides. Ideally, these nucleotides are incorporated specifically, cleaved efficiently during or following fluorescence imaging, and extended as modified or natural bases in ensuing cycles. In the figure, red chemical structures denote terminating functional groups, except in the Helicos BioSciences structure, which is characterized by an inhibitory function[33]. Arrows indicate the site of cleavage separating the fluorophore from the nucleotide, and the blue chemical structures denote residual linker structures or molecular scars that are attached to the base and accumulate with subsequent cycles. DNA synthesis is terminated by reversible terminators following the incorporation of one modified nucleotide by DNA polymerase. Two types of reversible terminators have been described: 3′-blocked terminators, which contain a cleavable group attached to the 3′-oxygen of the 2′-deoxyribose sugar, and 3′-unblocked terminators.

Several blocking groups have been described (see the figure, part **a**), including 3′-*O*-allyl[19,21,101] (Ju & colleagues, who exclusively licensed their technology to Intelligent Bio-Systems) and 3′-*O*-azidomethyl[22,23,101] (Illumina/Solexa). The blocking group attached to the 3′ end causes a bias against incorporation with DNA polymerase. Mutagenesis of DNA polymerase is required to facilitate the incorporation of 3′-blocked terminators.

3′-unblocked reversible terminators (part **b**) show more favourable enzymatic incorporation and, in some cases, can be incorporated as well as a natural nucleotide using wild-type DNA polymerases[31]. Other groups, including Church and colleagues[102] and Turcatti and colleagues[103], have described 3′-unblocked terminators that rely on steric hindrance of the bulky dye group to inhibit incorporation after the addition of the first nucleotide.

With real-time nucleotides (part **c**), the fluorophore is attached to the terminal phosphate group (Life/VisiGen[16] and Pacific Biosciences[15]) rather than the nucleobase, which also reduces bias against incorporation with DNA polymerase. In addition to labelling the terminal phosphate group, LI-COR Biosciences' nucleotides attach a quencher molecule to the base[17]. Gamma-labelled 2′-deoxyribonucleoside triphosphates (dNTPs) were first described in 1979 by Yarbrough *et al.*[104], and more recently, Kumar *et al.* described their terminally labelled polyphosphate nucleotides[105]. With the exception of LI-COR Biosciences' nucleotides, which leave the quencher group attached, natural bases are incorporated into the growing primer strand.

©2010 Macmillan Publishers Limited. All rights reserved

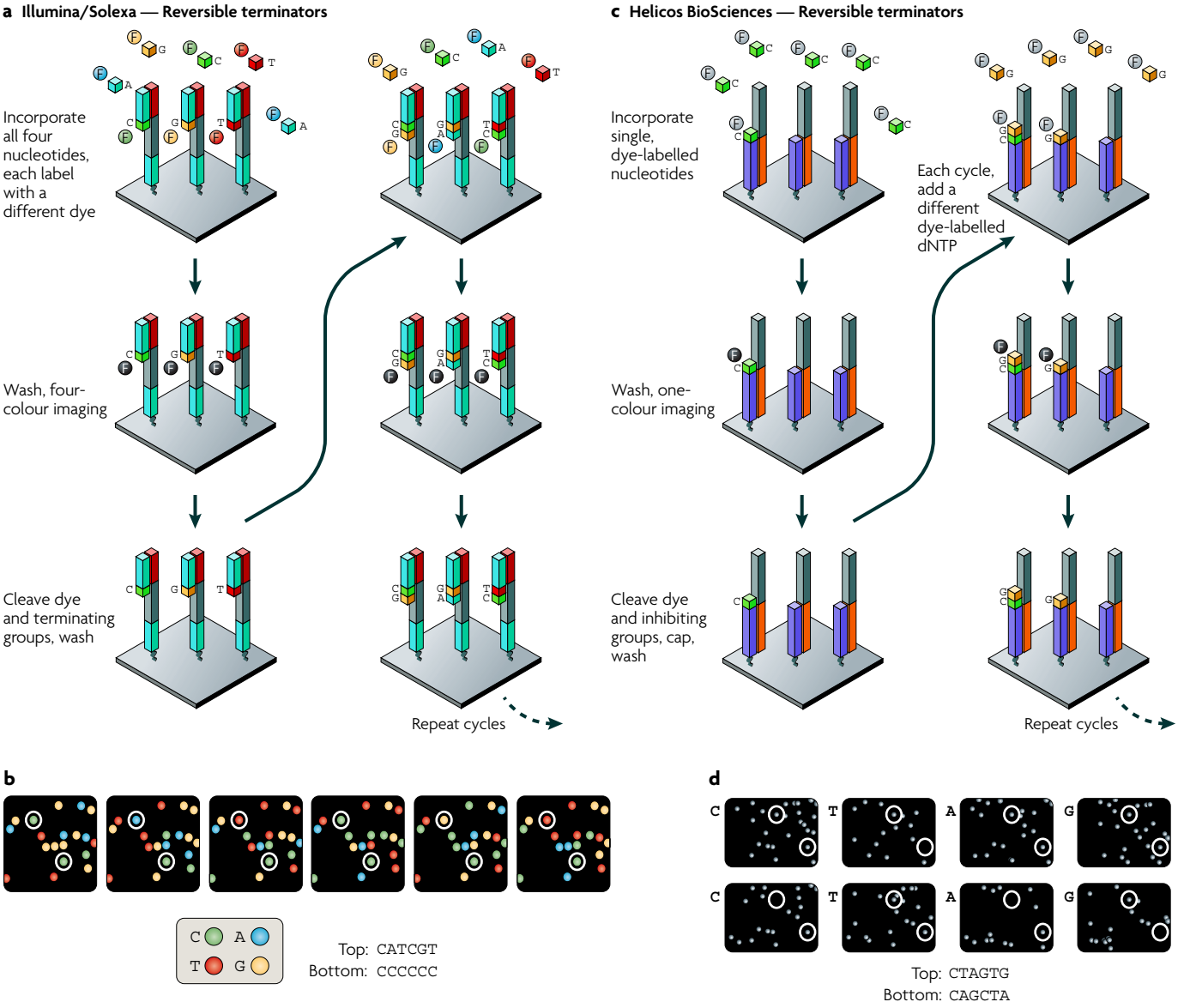


Figure 2 | **Four-colour and one-colour cyclic reversible termination methods. a** | The four-colour cyclic reversible termination (CRT) method uses Illumina/Solexa's 3′-*O*-azidomethyl reversible terminator chemistry[23,101] (BOX 1) using solid-phase-amplified template clusters (FIG. 1b, shown as single templates for illustrative purposes). Following imaging, a cleavage step removes the fluorescent dyes and regenerates the 3′-OH group using the reducing agent tris(2-carboxyethyl)phosphine (TCEP)[23]. **b** | The four-colour images highlight the sequencing data from two clonally amplified templates. **c** | Unlike Illumina/Solexa's terminators, the Helicos Virtual Terminators[33] are labelled with the same dye and dispensed individually in a predetermined order, analogous to a single-nucleotide addition method. Following total internal reflection fluorescence imaging, a cleavage step removes the fluorescent dye and inhibitory groups using TCEP to permit the addition of the next Cy5-2′-deoxyribonucleoside triphosphate (dNTP) analogue. The free sulphhydryl groups are then capped with iodoacetamide before the next nucleotide addition[33] (step not shown). **d** | The one-colour images highlight the sequencing data from two single-molecule templates.

One-base-encoded probe
An oligonucleotide sequence in which one interrogation base is associated with a particular dye (for example, A in the first position corresponds to a green dye). An example of a one-base degenerate probe set is '1-probes', which indicates that the first nucleotide is the interrogation base. The remaining bases consist of either degenerate (four possible bases) or universal bases.

Caenorhabditis elegans genome. From a single HeliScope run using only 7 of the instrument's 50 channels, approximately 2.8 Gb of high-quality data were generated in 8 days from >25-base consensus reads with 0, 1 or 2 errors. Greater than 99% coverage of the genome was reported, and for regions that showed >5-fold coverage, the consensus accuracy was 99.999% (J. W. Efcavitch, personal communication).

Sequencing by ligation. SBL is another cyclic method that differs from CRT in its use of DNA ligase[35] and either one-base-encoded probes or two-base-encoded probes. In its simplest form, a fluorescently labelled probe hybridizes to its complementary sequence adjacent to the primed template. DNA ligase is then added to join the dye-labelled probe to the primer. Non-ligated probes are washed away, followed by fluorescence

©2010 Macmillan Publishers Limited. All rights reserved

Table 1 | **Comparison of next-generation sequencing platforms**

Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US$)	Pros	Cons	Biological applications	Refs
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo-polymer repeats	Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics	D. Muzny, pers. comm.
Illumina/ Solexa's GA_{II}	Frag, MP/ solid-phase	RTs	75 or 100	4‡, 9§	18‡, 35§	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7‡, 14§	30‡, 50§	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Polonator G.007	MP only/ emPCR	Non-cleavable probe SBL	26	5§	12§	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery	J. Edwards, pers. comm.
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8‡	37‡	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods	91
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks	S. Turner, pers. comm.

*Average read-lengths. ‡Fragment run. §Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLiD, support oligonucleotide ligation detection.

imaging to determine the identity of the ligated probe[36]. The cycle can be repeated either by using cleavable probes to remove the fluorescent dye and regenerate a 5′-PO_4 group for subsequent ligation cycles (FIG. 3a) or by removing and hybridizing a new primer to the template (not shown in the figure).

Shendure and colleagues[11] used the SBL method to sequence the *Escherichia coli* MG1655 genome. Mate-pair templates were prepared with four priming sites (named A1 to A4) and were amplified by emPCR (FIG. 1a). A number of one-base-encoded probes, 1-probes to 7-probes, were used. In the first SBL cycle, the A1 primer was annealed to the template, followed by the hybridization and ligation of 1-probes, four-colour imaging and removal of the entire primer–probe strand from the solid-phase-bound template. The SBL cycle was then repeated with the A1 primer but using 2-probes, and then with the A1 primer but using 3-probes, and so on. The other three primers, A2, A3 and A4, were then cycled in an analogous manner to yield six (A2 and A4) and seven (A1 and A3) base reads (albeit discontiguous) from each genomic end, making a total of 26 base reads per mate-pair template. From two instrument runs, the authors reported the production of approximately 48 million high-quality bases, which mapped to about 70% of the *E. coli* genome[11]. This SBL method is being used on the Polonator instrument.

Life/APG has commercialized their SBL platform called support oligonucleotide ligation detection (SOLiD)[37]. The method uses two-base-encoded probes, which has the primary advantage of improved accuracy in colour calling and SNV calling, the latter of which requires an adjacent valid colour change. Colour space is a unique feature of the SOLiD system. A primer is

Two-base-encoded probe
An oligonucleotide sequence in which two interrogation bases are associated with a particular dye (for example, AA, CC, GG and TT are coded with a blue dye). '1,2-probes' indicates that the first and second nucleotides are the interrogation bases. The remaining bases consist of either degenerate or universal bases.

©2010 Macmillan Publishers Limited. All rights reserved