

By David Hertzog, Megan N. Hersh, and Michael L. Metzker

# A high-performance, low-cost approach to next-generation sequencing

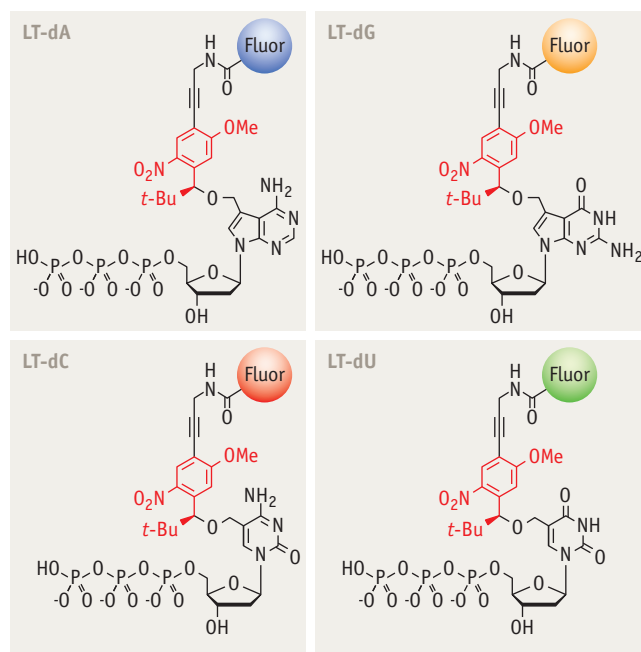
*A combination of chemistry advancements to boost accuracy of primary data, powerful software to facilitate speed and control, and flexibility to support multiple imaging approaches yields a robust, low-cost approach to DNA sequencing. The method promises to empower many applications beyond research—including personalized medicine.*

Next-generation sequencing (NGS) technologies have transformed scientific approaches in basic and applied research by enabling enormous volumes of DNA sequencing data at low cost.<sup>1</sup> The tradeoff to this advantage, however, is that NGS technologies generally produce primary read data of lower quality compared with those created by the gold standard, Sanger sequencing.<sup>2</sup>

Efforts to overcome this limitation are seldom reported in genome costs, yet they can be significant. Innovations in creating a novel nucleotide chemistry paradigm have resulted in improved accuracy of the primary data, which will inherently reduce genome cost. Speed of the new system is afforded by a combination of the performance of the novel chemistry and control software. Once commercialized, such a system could reach beyond the bounds of research to affect forensics, biodefense, pharmacogenomics, and diagnostics, and eventually enable personalized medicine.

## NGS error sources

Introduction of errors can occur during template preparation, sequencing and imaging, and/or read mapping to a reference genome. For stochastic errors, increasing the read coverage across any given nucleotide position can improve accuracy in the final genome sequence product. Unfortunately, this approach only partially offsets the cost advantages because it



**FIGURE 1.** In *Lightning Terminators* (LT), red chemical structures denote terminating functional groups that cleave when exposed to UV light.

DAVID HERTZOG, Ph.D., is VP of business development and MEGAN N. HERSH, Ph.D., is a staff scientist with LaserGen (Houston, TX; [www.lasergen.com](http://www.lasergen.com)). MICHAEL L. METZKER, Ph.D., is LaserGen's founder, president, and CEO, as well as associate professor of Molecular & Human Genetics and assistant director, Future Technologies of the Human Genome Sequencing Center at Baylor College of Medicine (Houston, TX; [www.bcm.edu/genetics/index.cfm?pmid=10947](http://www.bcm.edu/genetics/index.cfm?pmid=10947)). Contact Dr. Metzker at [mmetzker@lasergen.com](mailto:mmetzker@lasergen.com).

requires more reads from a given instrument run. This reduces the overall utilization of read production per genome, while requiring a substantial increase in information technology (IT) storage and data processing. In some cases, IT costs can outweigh that of the NGS read production itself.<sup>3</sup>

Systematic errors are more challenging to detect, and although reports of important biomedical discoveries have identified molecular etiologies, these discoveries were facilitated by significant efforts to validate sequence variants using orthogonal technologies.<sup>4-7</sup> Validation efforts are important for reducing false positives, yet the current depth of coverage reported in whole genomes analyzed by NGS technologies may still be insufficient to catalog important sequence variants.<sup>8</sup>

### Goal: improved accuracy

Focusing on NGS chemistry improvements in order to produce more accurate primary read data, we aim to reduce both the IT burden and validation requirement. Our chemistry efforts have led to a novel 3'-OH unblocked reversible terminator that exhibits such properties as single-base termination, fast incorporation, high nucleotide selectivity, and rapid terminating group cleavage (see Fig. 1).<sup>9-12</sup>

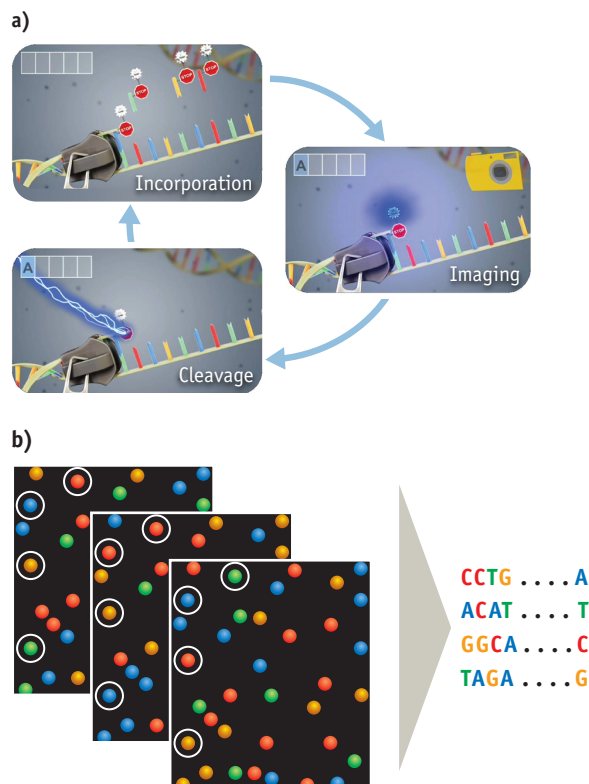
The reversible terminator chemistry, *Lightning Terminators*, has the potential to improve accuracy and read lengths significantly.<sup>1,2</sup> The cyclic reversible termination (CRT) method comprises three steps (see Fig. 2a):

- 1.) Incorporating reversible terminators using DNA polymerase (depicted as a zipper),
- 2.) Imaging the fluorescently labeled DNA molecules, and
- 3.) Photochemically cleaving the terminating and fluorescent groups with ultraviolet (UV) light to restore the modified nucleic acid to its native state.

Base calling is then performed from processed fluorescent intensities of individual beads. The CRT method's read length is a direct function of the number of cycles it executes (see Fig. 2b).

To showcase these chemistry improvements, we have partnered with National Instruments (Austin, TX) to assist LaserGen in building its early stage prototype system that will be tested against other NGS technologies for sequencing of the *E. coli* genome.<sup>13</sup> In early 2012, LaserGen's commercialization phase of development will begin with place-

**FIGURE 2.** a) A single DNA molecule serves to diagram the cyclic reversible termination (CRT) method's three steps: incorporation, fluorescence imaging, and photochemical cleavage; and b) four-color tile images from three cycles and subsequent base calling from individual beads.

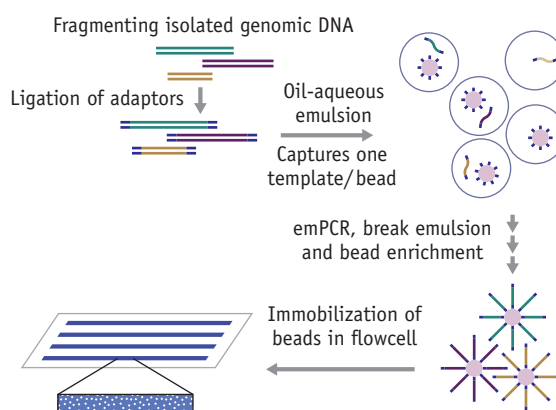


ment of the prototype system in a leading U.S. genome center.

### Workflow and instrumentation

Two primary methods used to prepare DNA of interest for NGS analysis are clonally amplified and non-amplified (i.e., single molecule) templates.<sup>1</sup> While *Lightning Terminators* can be formatted using either template preparation strategy, our initial system will employ the clonally amplified method of emulsion PCR (emPCR).<sup>14</sup>

Our front-end workflow involves isolating genomic DNA, fragmenting it into smaller pieces, and ligating common adaptors to the ends of those fragments (see Fig. 3). We then separate adaptor-ligated DNA molecules into single strands and associate them with 1  $\mu$ m size beads under conditions that favor one DNA molecule per bead. An oil-aqueous emulsion creates individual droplets that encapsulate these bead-DNA complexes. We perform PCR amplification within these droplets to create beads containing  $10^4$ – $10^6$  copies of the



**FIGURE 3.** Front-end preparation of clonally amplified templates immobilized in a flowcell. Emulsion PCR (emPCR) is performed on single DNA molecules that have common adaptors ligated to the ends of fragmented genomic DNA. Template-positive emPCR-amplified beads are then immobilized within the 1-, 4-, or 8-channel flowcell (four channels are shown for illustration purposes only).

same template sequence.

Following successful amplification and enrichment, we then chemically immobilize tens to hundreds of millions of emPCR beads to a custom-designed, 1-, 4-, or 8-channel glass flowcell, which is then installed into the system.

The chosen template preparation method directly impacts selection of imaging technologies, and given the chemistry's flexibility in adapting to different template formats, we can employ a wide range of imaging approaches, including standard four-color epi-fluorescence, total internal reflection fluorescence (TIRF), or LaserGen's color-blind, pulsed-multiline excitation (PME) approach.<sup>15, 16</sup> The prototype system incorporates many 'off-the-shelf' components. In the optical path, a four-color, epi-fluorescence microscopy configuration uses a digital camera to capture fluorescent signals derived from emPCR-amplified template beads immobilized in the microfluidic flowcell. Each channel is imaged in a tiled fashion to collect a majority of the fluorescent signals within the flowcell area; a motorized filter wheel switches the appropriate spectral filters at each position. Next, a UV LED light source provides broad illumination across the flowcell for the photochemical cleavage step. The terminating and fluorescent dye groups are then washed away to begin the next cycle (see Fig. 4).

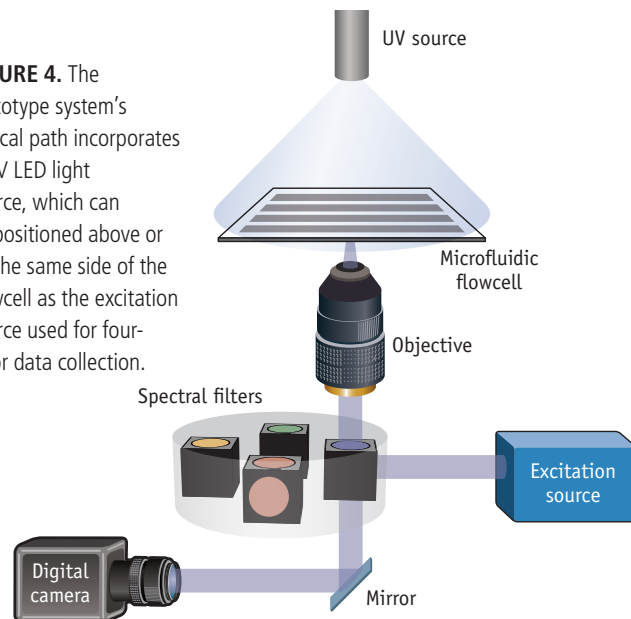
Particle finding and alignment, base calling, system control, and data analysis are all enabled by custom software and proprietary algorithms governed by National Instruments' virtual instrumentation software, LabVIEW. This software tool governs all steps required for the CRT method, including the addition of the *Lightning Terminators* cocktail, washes to remove non-reacted reagents, four-color imaging, and photochemical cleavage, and washes to remove the terminating and fluorescent groups. LabVIEW software's performance and speed of both system control and data analysis allow for small reagent volumes, cycle times of less than 30 minutes, real-time data analysis, and throughputs of up to five gigabases per day.

### Development and applications

As faster cycles result in shorter turnaround times for delivery of genome sequences, a goal of the instrument design is integration of a hardware/software infrastructure that approaches the fast rates of incorporation and photochemical cleavage of the *Lightning Terminators* reagent set.<sup>11,12</sup> We anticipate the prototype system being cost-competitive with other NGS platforms, such as Illumina's MiSeq and Ion Torrent's Personal Genome Machine, while offering the key advantage of delivering more accurate primary data—with a goal of Sanger-quality reads.

The system's potential sequencing applications include pathogen genomes, targeted capture of whole exome and regional genomic loci, and seq-based methods such as RNA-seq, ChIP-seq, and methyl-seq, among others. Beyond the research market, the system is expected to serve forensics, biodefense, pharmacogenomics, diagnostics, and eventually personal medicine, all of which are fields that demand accurate sequence infor-

**FIGURE 4.** The prototype system's optical path incorporates a UV LED light source, which can be positioned above or on the same side of the flowcell as the excitation source used for four-color data collection.



mation. The flexibility offered by compatibility with different template and imaging platforms could expand applications into the complete sequencing of mammalian genomes or quantitative analysis, resulting in digital readouts for diagnostic purposes. <<

### ACKNOWLEDGMENTS

The authors thank Sherry Metzker and Shane Climie for critical reading of the manuscript.

*Lightning Terminators* is a trademark of LaserGen, and LabVIEW is a trademark of National Instruments.

### REFERENCES

1. M. L. Metzker, *Nature Rev. Genet.* 11, 31–46 (2010).
2. M. L. Metzker, *Genome Res.* 15, 1767–1776 (2005).
3. L. Stein, *Genome Biol.* 11, 207 (2010).
4. K. Nakamura et al., *Nucleic Acids Res.* 39, e90 (2011).
5. J. R. Lupski et al., *N. Engl. J. Med.* 362, 1181–1191 (2010).
6. M. N. Bainbridge et al., *Sci. Trans. Med.* 3, 87re83 (2011).
7. N. Agrawal et al., *Science* 333, 1154–1157 (2011).
8. J. Gray, *Nature* 464, 989–990 (2010).
9. W. Wu et al., *Nucleic Acid Res.* 35, 6339–6349 (2007).
10. V. A. Litosh et al., *Nucleic Acid Res.* 39, e39 (2011).
11. B. P. Stupi et al., "Stereochemistry of benzylic carbon substitution coupled with ring modification of 2-nitrobenzyl groups as key determinants for fast-cleaving reversible terminators," submitted for publication (2011).
12. M. N. Hersh et al., unpublished data (2011).
13. <http://www.globenewswire.com/newsroom/newshtml?ref=rss&d=226127>.
14. D. Dressman et al., *PNAS USA* 100, 8817–8822 (2003).
15. D. Axelrod, *J. Cell Biol.* 89, 141–145 (1981).
16. E. K. Lewis et al., *PNAS USA* 102, 5346–5351 (2005).