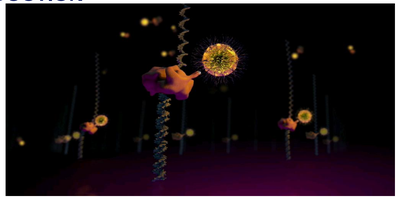
Single Molecule Real-time DNA Sequencing using FRET-based reagents: sequencing DNA on multiple size scales (from single bases to whole chromosomes) to resolve structural variation and enable de novo sequencing.

Joseph M. Beechem, Umberto Ulmanello, YuFang Wang, Min Yue, Mike Lafferty and Hong Ye Sun, Genetic Systems, Life Technologies, 1620 Faraday Ave, Carlsbad CA 92008 (joe.beechem@lifetech.com)

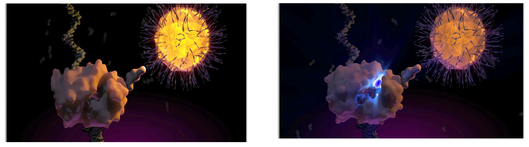
ABSTRACT

A single molecule, long read-length, real-time sequencing-by-synthesis technology has been developed by building a sequencer directly onto nanometer-sized particles. Fluorescence resonance energy-transfer (FRET), which operates over distance scales of approximately 10 nanometers, supplies all of the needed signal localization (molecularly) to perform high resolution single molecule studies. Up to five-color fluorescence resonance energy-transfer technology (FRET) is utilized for DNA sequence detection; signals from the nanoparticle-labeled DNA polymerase plus 4 DNA-base-specific acceptor dyes are simultaneously detected. Dye attachment to the nucleotide is via the terminal phosphate, released into solution after base-incorporation, yielding natural DNA during synthesis. Because the sequencer is not physically bound to any solid substrate, it can be exchanged (like a reagent) during mid-sequence runs, replacing damaged non-functioning polymerases mid-reaction, greatly extending net read-length capability. In a new demonstration of reagent-based sequencing flexibility, these sequencers can bind to ultra-long DNA segments at multiple positions and can sequence DNA while moving "horizontally" (parallel to TIRF field) along DNA that is "laying-down". In this "Top-Down" sequencing method, one can obtain large-scale structural variation information by imaging sequencing (or related) nanoparticle distribution along the long-DNA, and more detailed single-base-resolution information is obtained as the run progresses. This technology is still in the research-feasibility stage of development. Multiple single-molecule sequencers bound to individual elongated 4kb, 8kb, 50kb, and 150kb fragments will be shown and sequence-dependent chemical mapping and sequencing demonstrated. We believe this top-down sequencing approach provides an ideal method to directly measure how long-range structural features in DNA/RNA map (or any element of long-DNA that can be spatially imaged) to localized sequence and biological function.

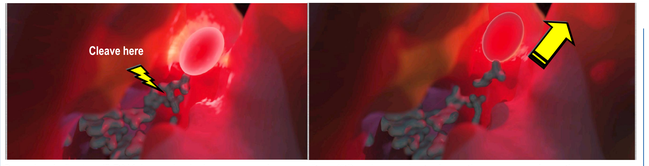
INTRODUCTION



Nanometer-sized sequencing machine. Genomic DNA is chemically attached to the coverslip of a standard TIRF-based (Total Internal Reflection) microscope. Universal primers are annealed to the genomic DNA. Nanoparticle-labeled DNA polymerase is added to the slide, and binds to the primer-template end. In the final step, the sequencing-by-synthesis reaction is initiated by the addition of 4 terminal-phosphate-dye labeled nucleotides.



As the polymerase binds the incoming terminally-phosphate-dye-labeled nucleotide (different colored dye for each base, ATC or G), there is fluorescence resonance energy-transfer (FRET) as photons resonantly couple from the nanoparticle to the incoming dye (right), decreasing the photon-flux out of the nanoparticle and generating a photon-flux in the base-specific (ATCG) dye-color.

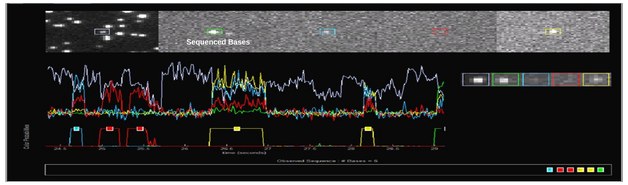


Completely "natural" DNA is made during sequencing process, as the dye is connected to the deoxynucleotide leaving-group (cleavage at alpha-beta phosphate bond, dye off terminal phosphate).

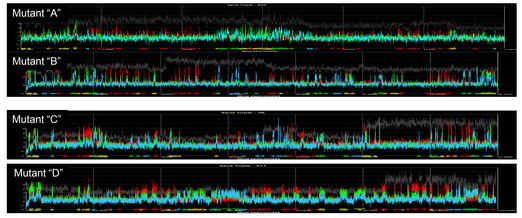
Dye-(Pi)_n by-product of reaction escapes from the polymerase binding channel (yellow arrow) as DNA is synthesized. Donor and Acceptor signals return to baseline, the polymerase moves to the next base to incorporate, and the entire sequence of events repeats.

Thousands of individual single-molecule sequencing machines are imaged simultaneously in real-time. Each individual sequencer generates a time-series data stream representing the (originally unknown) sequence of the genomic DNA originally captured in step (1) of the reaction sequence (first figure above).

Each individual sequencer generates a 5-color time-series, with nanoparticle Donor-photon-drop signals for each base incorporated (yellow trace) and nucleotide-dye acceptor photon-increases with a unique color associated with each base type (T = blue, A = red, G = green, C = purple). Time-scale for a typical incorporation event ~ 50 milliseconds.

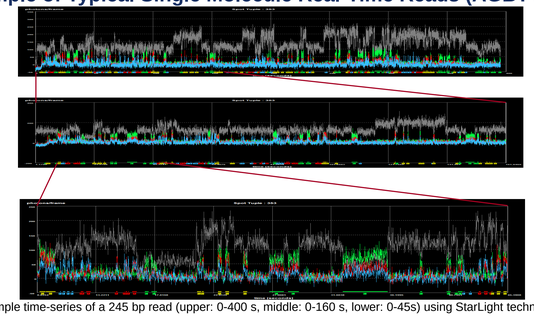


Detailed view of a quantum-dot 5-color real-time single molecule sequencing reaction (expanded view of 4-seconds of data). (Upper) small segment of the field-of-view of the DNA sequencer containing ~ 20 nanosequencers imaged at 5 distinct colors (donor color #1, plus the 4 base (GATC) colors). (Middle left): Time-series associated with a single nanosequencer highlighted in upper left. (Lower) Real-time color-to-GATC-base calling of the time-series data.



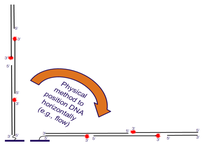
Progressive development and advancement of the StarLight sequencing engine via mutational analysis of various DNA polymerases ("A" = early generation (AGBT 2010) to "D" latest generation (AGBT 2011)).

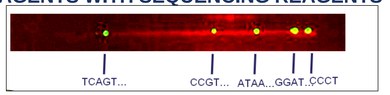
Example of Typical Single Molecule Real-Time Reads (AGBT 2011)



Example time-series of a 245 bp read (upper: 0-400 s, middle: 0-160 s, lower: 0-45s) using StarLight technology.

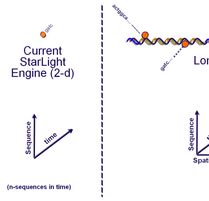
3-DIMENSIONAL DNA SEQUENCING: COMBINING IMAGING REAGENTS WITH SEQUENCING REAGENTS

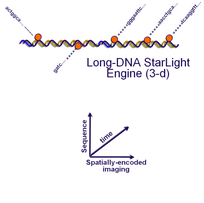


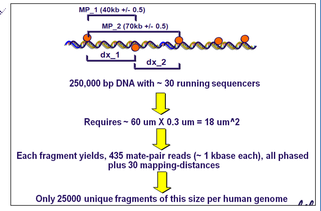


48.5 KB

Single-frame image of approximately five (visible) nanosequencers (colored-green) bound to a single 48.5 Kilobase strand of nicked/gapped-DNA (lightly labeled with SYTOX Orange to make the double-stranded DNA visible). Being able to simultaneously sequence from multiple sites of a single DNA molecule allow for greatly enhanced DNA-phasing, long-range scaffold generation, large-repeat resolution, as well as many additional sequencing advantages (see below).

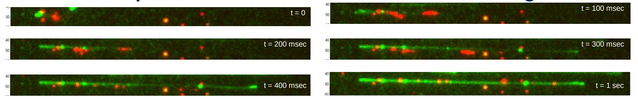






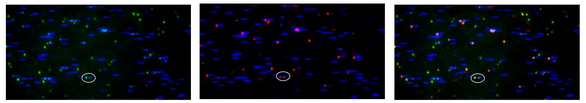
(Left) Real-time sequencing generates exceptional high information content 2-D data (sequence vs time). However, when performing real-time sequencing on DNA fragments large enough to spatially image (> a few kilobases), an entirely new dimension of information can be obtained (sequencing vs time vs imaging-reagent). This added dimension of information allows one to directly (in a single experiment) measure correlations between local DNA sequence and any structural element of DNA that can be spatially imaged (e.g, restriction sites, methylation sites, promoter-sites, etc).

Nano-sequencers bound to 144 Kilobase DNA Fragments

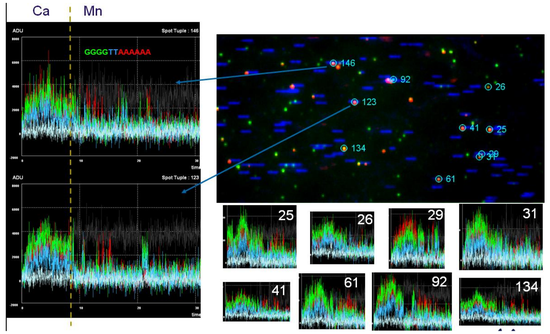


Time-series of multiple nanosequencers (red and green "dots") bound along the length of a single144 Kilobase length of (predominantly) dsDNA (Sytox Green image) as the DNA is elongated in sequencing flow chamber.

SEQUENCING ON DEFINED 8.4 Kilobase DNA FRAGMENTS



Multiple nanosequencers bound along the length of 8 Kilobase defined, gapped primer-template. DNA imaged (with DAPI) is blue. (Left) DNA (blue) and bound nanosequencers (green). (Middle) DNA (blue) and next correct deoxynucleotide (red). (Right) Superposition of (Left) and (Center), revealing sequencing-competent nanosequencers on the DNA. Note: some DNA's no longer visible (with DAPI) at final image frame. In white-circled-area one has 2-bound nanosequencers, but only the left-hand sequencer is sequencing competent.



Multiple nanosequencers bound along the length of 8 Kilobase defined, gapped primer-template. By using various salt conditions one can "freeze" the competent nanosequencers into binding the correct base, but not advancing (calcium-bound state). Exchange of Ca2+ with Mn2+ initiates real-time sequencing and multiple DNA fragments (labeled with number-identifiers) are followed as a function of time.

CONCLUSIONS

Life Technologies Single Molecule Real-Team Sequencing Technology (aka "StarLight") continues to advance in capability. The "standard" single molecule sequencing runs continue to advance at the DNA polymerase sequencing engine level (see column 2 lower) and at the integrated sequencing performance level (see column 3 upper). Especially new (on the horizon) is the ability to perform 3-Dimensional DNA sequencing of ultra-long DNA fragments, wherein DNA-sequence vs. time vs. imaging-reagent-space are simultaneously collected (see column 3 lower and column 4). This additional information provides the ability to simultaneously measure how sequencing correlates with any factor on DNA that can be spatially imaged (e.g., methylation, restriction sites, promoter sites, etc.). In addition, completely phased and ordered reads are simultaneously obtained, and the effective "mate-pairs" for each DNA fragment increase combinatorially with the number of sequencers on each individual DNA fragment. This type of 3-D sequencing information is ideal for quantitating genomic structural variation and for generating de novo scaffolds for shorter read-length Gen-2 sequencing data.

ACKNOWLEDGEMENTS

The entire Life Technologies StarLight sequencing staff at the: Carlsbad CA, Foster City CA, and Eugene OR sites. Special acknowledgement to the pioneering work of Dr. David Schwartz in the area of optical-mapping of large DNA fragments.