Application Note:
DNA Sequencing

Explore the Possibilities

With the potential to generate a billion bases of DNA sequence per run, the Solexa Genome Analysis System provides researchers with the opportunity to sequence mammalian genomes in a matter of weeks rather than years, as is the case with present technologies. Furthermore, since our Clonal Single Molecule Array technology does not rely on expensive devices to achieve its unparalleled data density, the Solexa Sequencing-by-Synthesis approach is highly economical. Leveraging this technology, researchers can potentially resequence genomes for less than 1% of their current costs. With the capacity to accommodate up to 8 samples per flow cell, the Solexa Genome Analysis System can be tailored to the demands of your application. Whether your genome of interest is large or small, the Solexa Genome Analysis System is the ideal tool for your sequencing project.

About Solexa

Solexa, Inc. (NASDAQ: SLXA) is developing and commercializing a new genetic analysis instrument system based on our novel Clonal Single Molecule Array™ technology and proprietary reversible terminator chemistry. This new platform, the Solexa Genome Analysis System, is being designed to offer a powerful new approach to some of today's most important genetic analysis applications, including:

- Whole Genome and Candidate Region resequencing
- Genome Wide Expression profiling
- Small RNA identification and quantification
- Hybridization array data validation

To learn more about Solexa, our technology, and our products, please visit www.solexa.com or email us at info@solexa.com

For more information on our Genomic Services, please contact us at:

Solexa Inc.
Corporate Headquarters

25861 Industrial Boulevard
Hayward, CA 94545
U.S.A.

P: 510-670-9300
F: 510-670-9302

info@solexa.com
www.solexa.com

Solexa Ltd.
European Headquarters

Chesterford Research Park
Little Chesterford
Essex, CB10 1XL
United Kingdom

P: +44 (0) 1799 532300
F: +44 (0) 1799 532301

The Solexa Genome Analysis System leverages the company's proprietary Clonal Single Molecule Array™ technology and reversible terminator chemistry for generating rapid, economical, and accurate DNA sequence data. With the potential to yield up to 1 billion bases of DNA sequence per run at less than 1% of the cost of capillary-based methods, the Solexa Genome Analysis System is designed to enable researchers to dramatically improve the efficiency and speed of current applications and undertake new research that has been limited by lack of suitable technologies.

THE WORKFLOW

- **Random fragmentation of genomic DNA**
- **Addition of short adapters to the DNA fragments**
- **Immobilization of modified fragments to the flow cell**
- **Solid phase amplification to generate millions of distinct DNA clusters**
- **Base-by-base sequencing-by-synthesis using fluorescent reversible terminators**
- **Base calling and QC**
- **Application specific analysis**

Clonal Single Molecule Array™ Technology

Sequencing templates are immobilized on a proprietary flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template and low non-specific binding of fluorescently labeled nucleotides. Solid phase amplification is employed to create up to 1,000 identical copies of each single molecule in close proximity (diameter of 1 micron or less). Since the process does not involve photolithography, mechanical spotting or positioning of beads into wells, the Clonal Single Molecule Array technology can achieve densities of up to 10 million single molecule clusters per square centimeter.

Sequencing-by-Synthesis

Solexa's Sequencing-By-Synthesis (SBS) utilizes four proprietary fluorescently labeled modified nucleotides to sequence the millions of DNA clusters present on the flow cell surface. These nucleotides, specially designed to possess a reversible termination property, allow each cycle of the sequencing reaction to occur simultaneously in the presence of all four nucleotides (A, C, T, G). In each cycle, the polymerase is able to select the correct base to incorporate, with the natural competition between all four alternatives leading to higher accuracy than methods where only one nucleotide is present in the reaction mix at a time. Sequences where a particular base is repeated one after another (e.g., homopolymers) are addressed like any other sequence and with high accuracy.

Analysis Pipeline

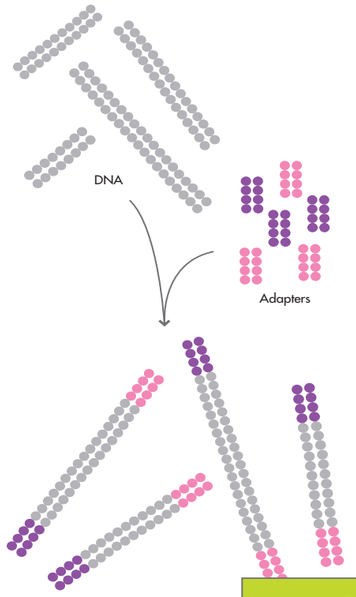
The Solexa sequencing approach is built around a very large number of short sequence reads. Deep sampling of more than ten-fold even coverage is required to generate a consensus and thus ensure high confidence in determination of genetic differences. Such differences are identified by comparison of sequence reads to a reference. Deep sampling allows the use of weighted "majority voting" and statistical analysis, similar to conventional methods, to identify homozygotes and heterozygotes and to distinguish sequencing errors. Each raw read base has an assigned quality score so that the software can apply a weighting factor in calling differences and generating confidence scores.

The suite of software from Solexa will enable users to align sequences to a reference in resequencing applications. Developed in collaboration with leading researchers, Solexa's software suite includes the full range of data collection, processing, and analysis modules to streamline collection and analysis of data with minimal user intervention. The open format of the software allows for easy access to the data at various stages of processing and analysis using simple application program interfaces.

Sequencing Technology Overview

1 Prepare genomic DNA sample

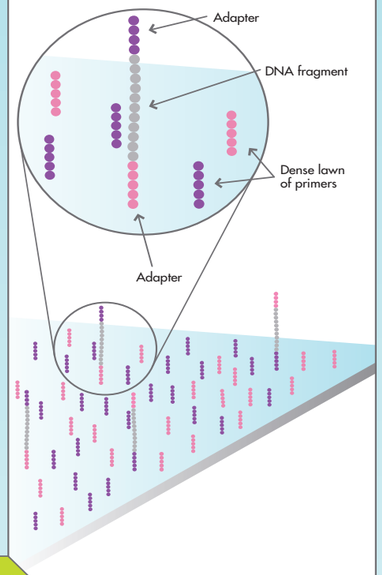
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.



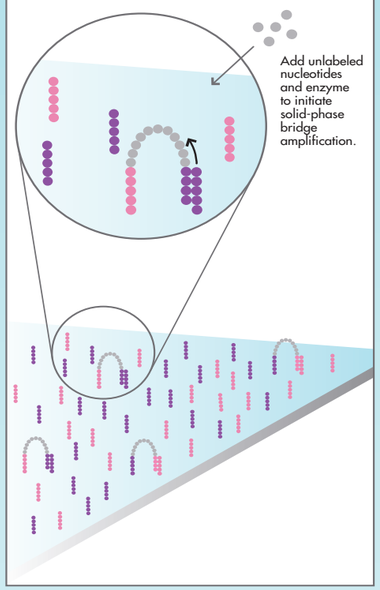
Add sample to flow cell

2 Attach DNA to surface

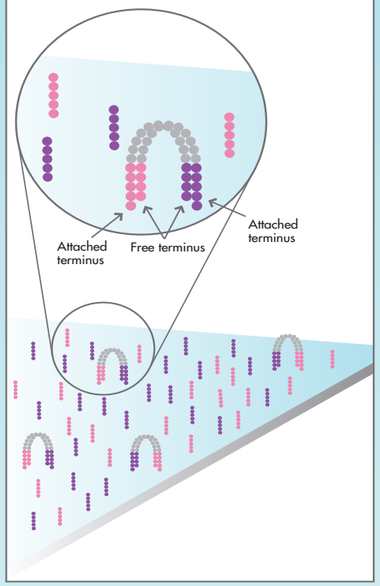
Bind single stranded fragments randomly to the inside surface of the flow cell channels.



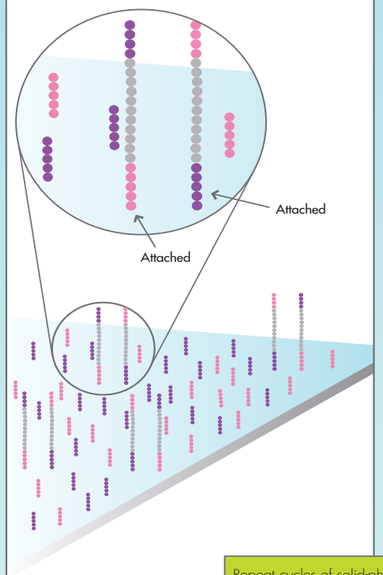
3 Bridge amplification



4 Fragments become double stranded



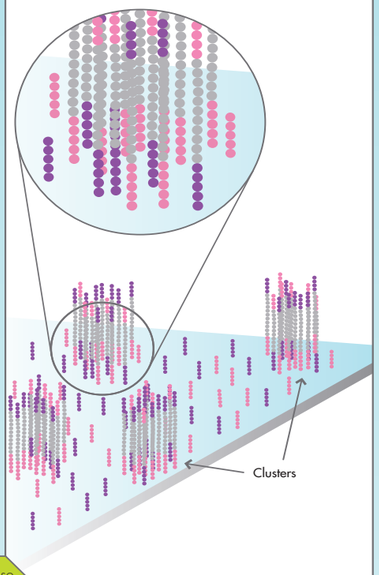
5 Denature the double stranded molecules



Repeat cycles of solid-phase bridge amplification

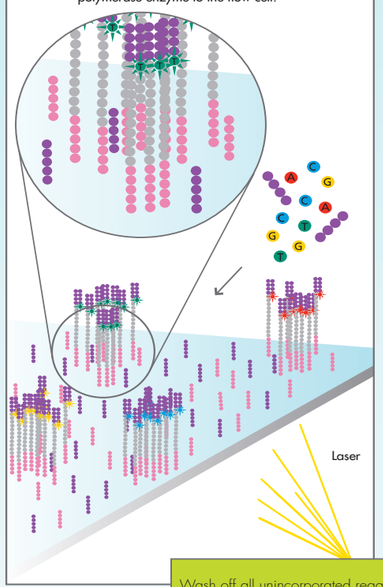
6 Completion of amplification

On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.



7 First chemistry cycle: determine first base

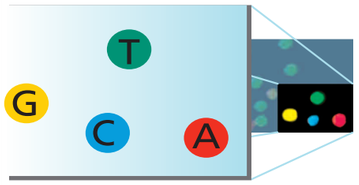
To initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.



Wash off all unincorporated reagents

8 Image of first chemistry cycle

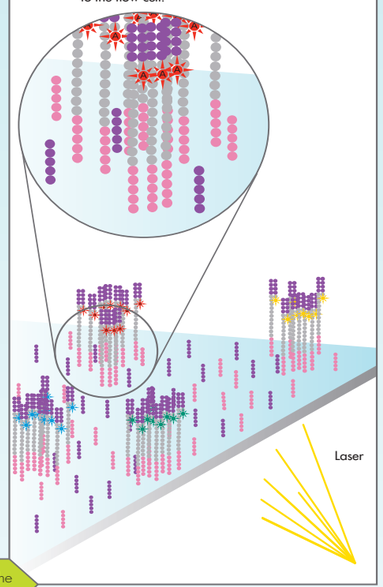
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.



Remove the blocked 3′ terminus and the fluorophore from each incorporated base

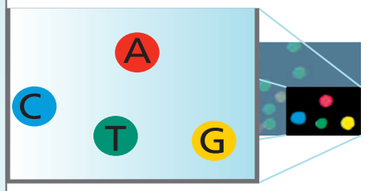
9 Second chemistry cycle: determine second base

To initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.



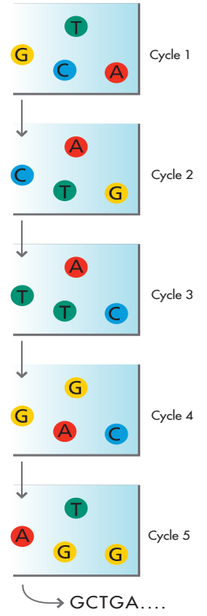
10 Image of second chemistry cycle is captured by the instrument

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.



11 Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.



12 Align the new data to a reference and identify sequence differences