Improvements to the Illumina Sequencing System and New Applications

David W. Hillman[1], Mary Ann Pedraza[2], Sirisha Sunkara[2], Maria Shin[1], Jeff Froula[2], Feng Chen[2] & Len Pennacchio[2]
[1]Lawrence Livermore National Laboratory, [2]Lawrence Berkeley National Laboratory

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Abstract

Next Generation sequencing products and procedures for the Illumina platform were evaluated, several of which immediately benefited JGI users. Paired end-reads of 35 bases each have become routine on our new GAii instruments with output in excess of 6 billion bases per run. Single reads have been used for polishing finished genomic sequence data for over 2 years, and collaborative work has progressed to include genomic SNP detection, expression analysis for eukaryotes and prokaryotes, and ChIP DNA sequencing. Paired End-Reads have been particularly useful for SNP detection. Our beta testing activities include positive results for barcode libraries (indexing) using Illumina's 3 primer scheme, large-gap libraries produced several ways, and accurate 75 base read-length. Key technologies include improvements to the instruments, reagents, software, and information management tools.

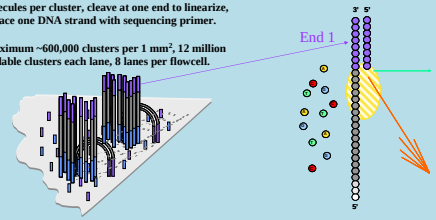
BASIC PRINCIPLE of the PROCESS

A. Make Clonal Single Molecule Arrays from shotgun DNA libraries.

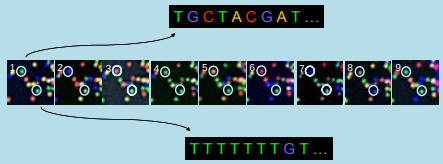
Hybridize to oligos on flowcell, amplify to ~1000 molecules per cluster, cleave at one end to linearize, replace one DNA strand with sequencing primer.

(Maximum ~600,000 clusters per 1 mm², 12 million readable clusters each lane, 8 lanes per flowcell.

B. Run 4-color Sequencing-By Synthesis, imaging each cycle.

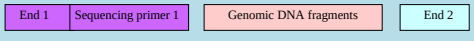


C. Call the bases by computer analysis of the series of photos.



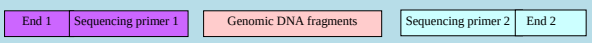
Shotgun Libraries

The process begins with genomic DNA fragmentation, ligation to dsDNA adapters, and size selection of the library. The ends are added during a 12 cycle PCR amplification.

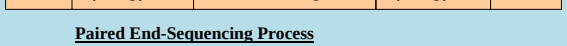


A Library of 120-200 bp insert clones is selected by agarose gel purification.

Paired-end libraries are made with a new PCR primer to add sequencing primer 2 to end 2.



A Library of 200-600 bp clones is selected by agarose gel purification.



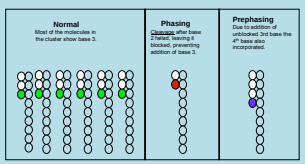
Paired End-Sequencing Process

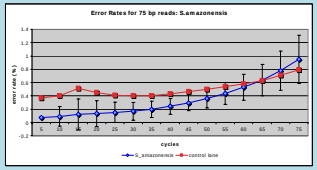
After completion of the single read process cluster amplification is repeated. Cleavage is done at the adapter 2 end (the opposite from read 1). A new sequencing primer for read 2 is hybridized, and the second read performed.

Recognition that both reads come from the same cluster requires that the coordinate system remain unchanged. This is accomplished by using the GAii fluidics with additional reagents from the Paired-End Module.

Illumina GAii System: New Hardware Components

LONGER READS USING IMPROVED CHEMISTRY





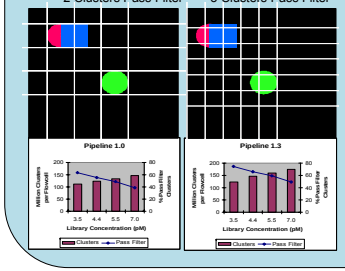
FASTER AND BETTER DATA PROCESSING

2007: Collect Data → Begin Image Analysis → Begin Basecalling
Present: Integrated Data Collection and Analysis (IPAR) → Finish Basecalling

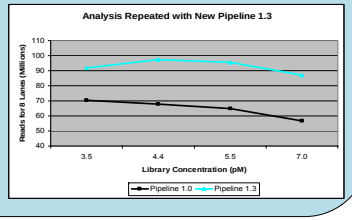
The IPAR Computer is placed below the GAii Instrument collecting data.

Pipeline 1.0
3 Clusters Detected
2 Clusters Pass Filter

Pipeline 1.3
3 Clusters Detected
3 Clusters Pass Filter

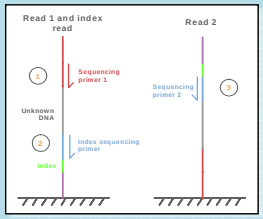


NEW ANALYSIS PROGRAM, Pipeline 1.3



NEW APPLICATIONS

Indexing, which allows mixing several libraries in one lane and separating reads base on a barcode.



Mixed-library lanes from 4 genomes, 3 barcodes each

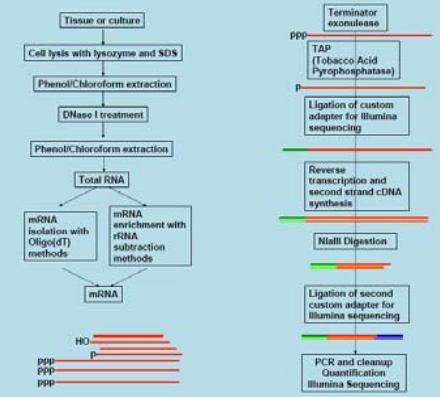
Plan

	lane	GBBC I-1	GBBC I-2	GBBC I-3	GBBF I-4	GBBF I-5	GBBF I-6	GBBH I-7	GBBH I-8	GBBH I-9	GBBI I-10	GBBI I-11	GBBI I-12
1%-10%	1	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
> 10%	2				33.3%	33.3%	33.3%						
<1%	3		25.0%			25.0%			25.0%		25.0%		
	4		12.5%			12.5%		12.5%	12.5%	12.5%	12.5%	12.5%	12.5%
	5		25.0%		25.0%			25.0%					25.0%
	6		12.5%		12.5%	12.5%	12.5%	12.5%			12.5%	12.5%	12.5%
	7	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
	8=C						100						

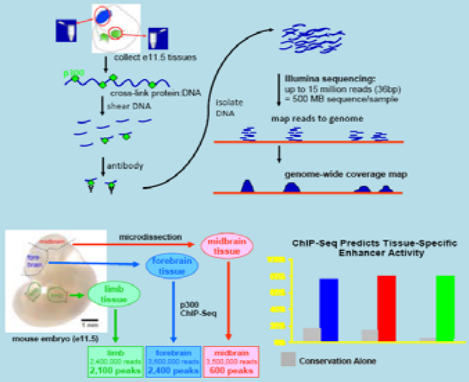
Data

	lane	GBBC I-1	GBBC I-2	GBBC I-3	GBBF I-4	GBBF I-5	GBBF I-6	GBBH I-7	GBBH I-8	GBBH I-9	GBBI I-10	GBBI I-11	GBBI I-12	Other
1%-10%	1	6.0%	6.3%	6.8%	6.7%	7.5%	8.4%	8.3%	7.7%	10.6%	9.3%	9.6%	8.5%	4.2%
> 10%	2				27.2%	32.1%	35.1%							5.7%
<1%	3		19.8%			23.8%			24.6%		28.3%			3.5%
	4		8.6%			10.7%		11.7%	11.9%	14.7%	13.1%	13.1%	12.7%	3.5%
	5		20.6%		21.2%			25.2%					27.9%	5.0%
	6		9.1%		9.7%	11.8%	12.8%	13.0%			13.2%	13.7%	13.1%	3.6%
	7	5.8%	6.0%	6.7%	6.7%	7.4%	8.3%	8.5%	7.9%	11.0%	9.5%	9.6%	8.6%	3.9%
	8=C							98.3%						1.7%

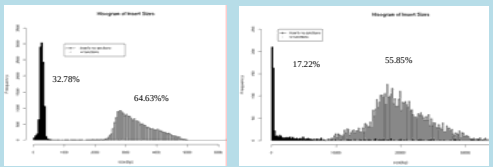
Transcription Start Site Sequencing



ChIP-Seq: Sequencing ChIP DNA on the Illumina platform



Large Gap Library Results



LLNL-POST-411717

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.