Illumina Production Sequencing at DOE Joint Genome Institute – Workflow and Optimizations

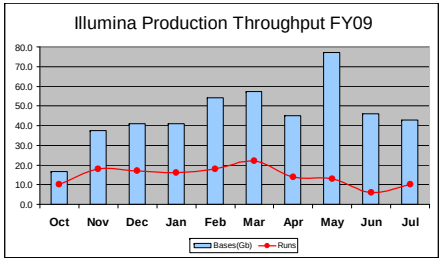
Angela Tarver[1], Matthew Zane[2], Christopher Daum[2], Mansi Chovatia[1], David Robinson[1], Christopher Hack[1], Feng Chen[1], Cindy Choi[1], Jan-Fang Cheng[1], Greg Mannino[1], Marty Pollard[1], Simon Roberts[1], Christine Naca[1], Melanie Alexandre[1], Susan Lucas[2]
[1] Lawrence Berkeley National Laboratory, [2]Lawrence Livermore National Laboratory

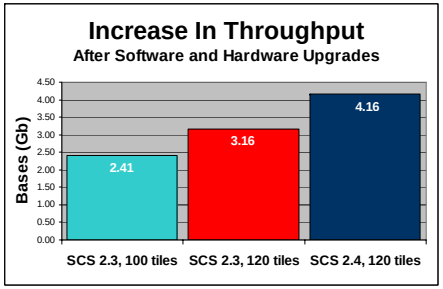
ABSTRACT The U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) Production Sequencing group is committed to the generation of high-quality genomic DNA sequence to support the DOE mission areas of renewable energy generation, global carbon management, and environmental characterization and clean-up. Within the JGI's Production Sequencing group, the Illumina Genome Analyzer pipeline has been established as one of three sequencing platforms, along with Roche/454 and ABI/Sanger. Optimization of the Illumina pipeline has been ongoing with the aim of continual process improvement of the laboratory workflow. These process improvement projects are being led by the JGI's Process Optimization, Sequencing Technologies, Instrumentation & Engineering, and the New Technology Production groups. Primary focus has been on improving the procedural ergonomics and the technicians' operating environment, reducing manually intensive technician operations with different tools, reducing associated production costs, and improving the overall process and generated sequence quality. The U.S. Department of Energy Joint Genome Institute (JGI) was established in 1997 in Walnut Creek, CA, to unite the expertise and resources of five national laboratories— Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Pacific Northwest – along with HudsonAlpha Institute for Biotechnology. JGI is operated by the University of California for the U.S. Department of Energy.

Production

In July 2008, Illumina transitioned to JGI production with a single GAII analyzer and two full time operators. By November 2008, the production group had acquired 2 more GAII analyzers, 4 paired-end modules, and 2 more cluster stations. In December 2008, a new LIMS system was released that included a run statistic summary report, an improved flow cell setup page with a reanalysis request page, a workflow for registering runs that will reduce run folder naming errors, and reliable script for uploading run data to the database. In June 2009, a new reporting layout was added to report basic run information. This is in addition to the html files generated by GERALD.

JGI Production Instruments		
5 GA*IIx* Analyzers	4 Paired-End Modules	3 Cluster Stations





FY09 Throughput Events:

November: Production purchased two new GAII's, four paired-end modules, and one cluster station

December: v3 Sequencing Kits

January: v2 Cluster Generation Kits

May: Conversion of genome analyzers from GAII to GAIIx format

June: Installation and utilization of Pipeline 1.4 and SCS 2.4 software

JGI Production FY09 Throughput to Date		
144 Runs	458 Gb	244 Projects

Scheduling Currently, a total of 3 to 4 staff is needed: 1 to 2 to generate libraries, and 2 to generate and load flow cells. Up to 16 libraries can be created a week by Cloning Technology, and each of the 2 Production operators are scheduled for two tasks daily.

	Monday	Tuesday	Wednesday	Thursday	Friday
Cloning Tech Operator 1	Library Creation Day 1 4 Libraries	Library Creation Day 2 4 Libraries	Library Creation Day 1 4 Libraries	Library Creation Day 2 4 Libraries	
Cloning Tech Operator 2	Library Creation Day 1 4 Libraries	Library Creation Day 2 4 Libraries	Library Creation Day 1 4 Libraries	Library Creation Day 2 4 Libraries	
Production Operator AM	Analyzer Run (PE) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries	Analyzer Run (PE) Up to 7 libraries Cluster Generation (SR) Up to 7 libraries	Analyzer Run (PE S/O) Up to 7 libraries Cluster Generation (SR) Up to 7 libraries	Analyzer Run (PE S/O) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries	Cluster Generation (PE) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries
Production Operator PM	Cluster Generation (SR) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries	Cluster Generation (SR) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries	Cluster Generation (SR) Up to 7 libraries Analyzer Run (PE S/O) Up to 7 libraries	Cluster Generation (SR) Up to 7 libraries Analyzer Run (PE S/O) Up to 7 libraries	Cluster Generation (PE) Up to 7 libraries Analyzer Run (SR) Up to 7 libraries

Ergonomics

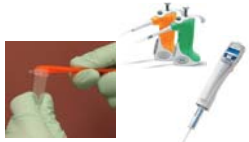
When the new technology laboratories were designed, features such as advanced ergonomic workstations, equipment, and tools that allow technicians to safely and efficiently prepare and sequence samples were incorporated.

Ergonomic lab work bench

Work benches Lab bench workspace is setup with height-adjustable tables and features cutouts. This recessed area is optimal for placement of supplies and arm/hand working heights.

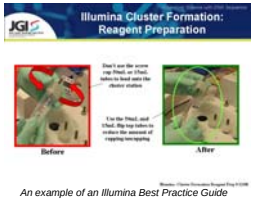
Bottle and Tube Openers There are a variety of tube and vial openers for operators to choose from.

Pipettes We have adopted the use of Ovation and XStream pipettes, in addition to the standard Rainin pipettes. These two alternatives reduce a "palm-up" position.



Ergonomic tube uncapping, and alternative pipettes

Best Practices In an effort to ensure that production work is performed safely and comfortably, required and best practice guides on how to perform lab work ergonomically were developed. An ergo points program was implemented to help determine the number and combination of work tasks that an operator can safely perform in a day.

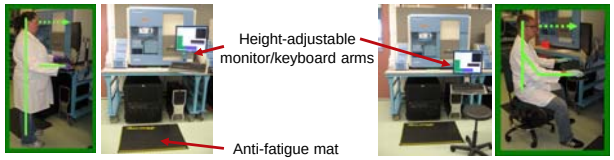


An example of an Illumina Best Practice Guide

Bottle Adapter Recently, the JGI instrumentation group developed a quick disconnect reagent bottle adapter for ease of loading and unloading reagents on the GA*IIx* analyzer. This bottle adapter helps the operator's hand, wrist and arm stay in a more neutral position.

Pictures of the quick disconnect bottle adapters for the Illumina GAIIx Analyzer

Analyzer Workstations The Illumina Analyzer workstations are setup with height-adjustable monitor/keyboard arms for operators of different height, and for ease of switching between a sitting and standing position. Each analyzer and cluster station also has an anti-fatigue mat to alleviate any discomfort when standing stationary for extended periods of time.

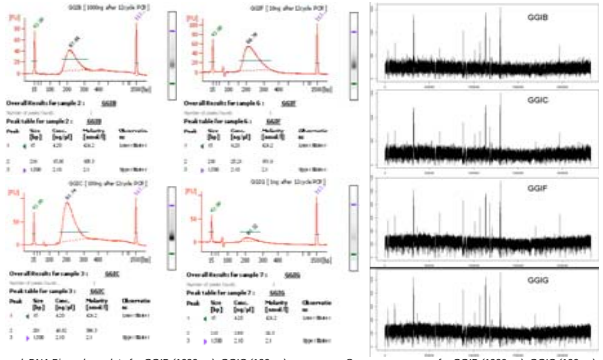


Standing position *Sitting position*

Developments

Trace Amount of DNA The amount of DNA provided for a given library is often very limited. Therefore, a protocol to minimize the amount of DNA needed for library construction for next-generation sequencing is very important. Our Cloning Technology group has begun to construct Illumina Paired-End libraries with less than 1 ug of genomic DNA. Two key components were changed in the protocol to minimize sample loss: shearing DNA via Covaris, and utilizing Agencourt® AMPure purification system.

Summary Table of Sequences and Genome Mapping Results						
Library Name	Sample Type	Number of reads used in the analysis	Average Size (bp)	Estimated Coverage	Aligned reads (%)	Gaps (%)
GGIB	1000ng	14,908,575	35	224.28	99.70%	1.08%
GGIC	100ng	14,908,575	35	223.58	99.39%	1.08%
GGIF	10ng	14,908,575	35	221.83	98.61%	1.08%
GGIG	1ng	14,908,575	35	216.21	96.11%	1.09%

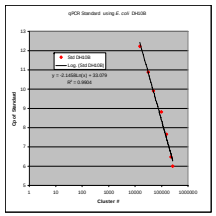


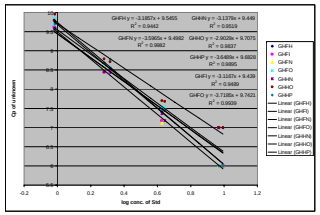
dsDNA Bioanalyzer data for GGIB (1000 ng), GGIC (100 ng), GGIF (10 ng) and GGIG (1 ng)

Sequence coverage for GGIB (1000 ng), GGIC (100 ng), GGIF (10 ng) and GGIG (1 ng)

The quality of sequence reads, assemblies and sequence coverage were very similar. We found a small decrease in the percent of total reads aligned to the genome when less starting material was used. We found striking similarity of the sequence depth across the genome.

Quantitative PCR Our Process Optimization group has been working on utilizing qPCR for estimating of a library's concentration needed for optimal cluster density. *E. coli* DH10B was used to create a standard curve and calculate the Cp (critical point). Currently, Production uses a mixture of titration runs and qPCR to extrapolate the optimal cluster density. The advantages of using qPCR are its sensitivity and precision used to generate the optimal number of clusters per lane, for each run.





LLNL-POST-415460

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.