Process Optimization and Assay Development for the Illumina Genome Analyzer Sequencing Platform

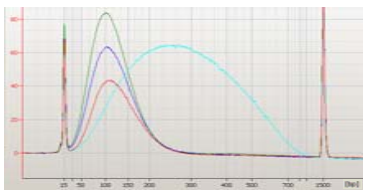
M. Costello, R. Hegarty, Z. Zwirko, I. Bourzgui, M. Lanzoni, N. Hafez, A. Barry, J. Meldrim, A. Gnirke,, R. Nicol, P. Cahill, S. Fisher
Genome Sequencing Platform
The Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141

Over the past year, we have been working on a variety of development projects aimed at improving the quality of our entire Illumina Genome Analyzer process:

- **Optimization of DNA shearing methods**
- **Development of a QPCR assay for library quantification**
- **Improvements to sample cleanup to eliminate excess enrichment PCR primer**
- **Implementation of an improved fluorescent microscopy flowcell QC method**
- **Development of fluorometric linearization and blocking QC assay**
- **Implementation of a method for rescuing failing flowcells on the Illumina GA sequencer**
- **Development of internal controls spiked into every lane on every flowcell**

Optimized Acoustic Shearing Method

Current shearing methods produce a broad distribution of fragment sizes from 50-700 bp. This is not optimal as the majority of DNA is not contained within the 125-175 bp range selected for in standard library size selection.

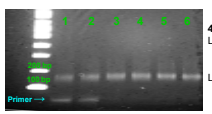


Agilent 1000 Chip results: The turquoise profile shows the previously validated acoustic shearing method using the Covaris E210 instrument. Blue, red, and green profiles show three samples treated with the improved method, which increased intensity and duty cycle.

Benefits of Improved Covaris E210 Shearing Method:
- Reduces sheared DNA size range from 50-700 bp to 20-200 bp
- Brings roughly 50% of our sheared DNA within the size range that we are cutting from the size selection gel
- Allows higher throughput shearing, up to 48 samples per set

SPRI for Post-Enrichment Cleanup

Excess enrichment primer in libraries can interfere with cluster amplification by binding to the grafted primers on the flowcell. Column based cleanup methods are inefficient at completely removing these enrichment primers..



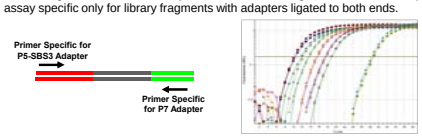
4-20% Polyacrylamide gel:
Lanes 1 - 2: Post MinElute PCR Cleanup
Lanes 3 - 6: Post SPRI PCR Cleanup

Implementation of a bead based cleanup using AMPure SPRI beads: SPRI is more efficient at removing the excess enrichment primer while still retaining the ~150 bp library band.

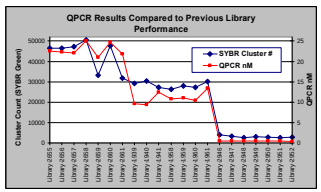
Benefits of SPRI Cleanup:
- Improved primer removal leads to easier quantification & more consistency
- Highly automatable

QPCR Assay for Library Quantification

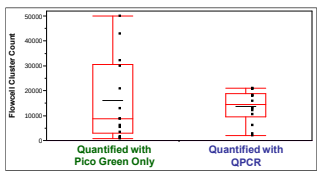
We currently quantify our libraries with Pico Green but see significant variation in downstream cluster number counts, likely caused by Pico Green binding to not only fragments with adapters at both ends but also fragments with one or no adapter and excess oligos. To overcome this problem, we have designed a SYBR Green QPCR assay specific only for library fragments with adapters ligated to both ends.



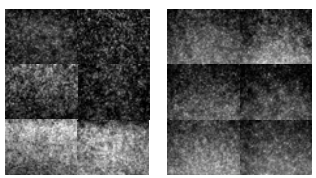
Overview of QPCR assay: At left, a simple schematic of the assay. Primers specific for each adapter amplify fragments in the presence of SYBR Green. Fluorescence is collected in real time on a Stratagene Mx3005P instrument. At right, QPCR amplification plots of a two-fold dilution series of PhiX control library.



Correlation of QPCR quantification results with library performance: A set of libraries previously cluster amplified based on Pico Green quantification results but with varied cluster number results were selected. When these libraries were re-quantified with QPCR, results correlated with the previously observed counts.



Reduction in cluster variability with QPCR quantification: The same set of libraries were loaded on flowcells based on both Pico Green and QPCR quant. The QPCR derived dilutions showed a decrease in variability of cluster counts.

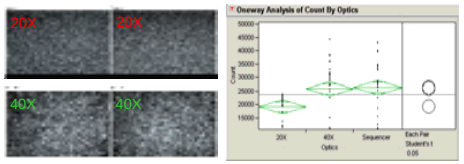


Flowcell comparison: At left, flowcell loaded at 4 pM based on Pico Green results. At right, flowcell loaded with same libraries at 4 pM based on QPCR results. Note that the QPCR flowcell has more consistent cluster counts between libraries.

Benefits of QPCR Quantification Method:
- Accurate prediction of cluster density for new libraries reduces variation, providing improved process stability and eliminates need to titrate
- Additional troubleshooting for problem libraries
- Reduces amount of rework necessary, therefore reducing excess cost

Fluorescent Microscopy Flowcell QC Assay

Failed runs on a sequencer due to poor quality flowcells take up valuable instrument time and lead to costly rework. A SYBR Green cluster amplification QC method has been implemented to catch problem flowcells before they reach the sequencers.

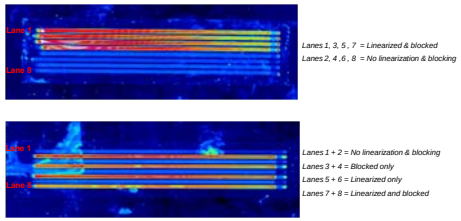


SYBR Green Flowcell QC Assay: At left, images of tiles from SYBR Green stained flowcell using both 20X and 40X objectives on a Olympus BX51 microscope . At right, comparing cluster counts obtained with the 20X and 40X objectives to those obtained by the sequencer. The 20X objective was originally implemented, but a change was made to 40X as these results correlate better with cluster counts from the sequencers.

Benefits of the 40X SYBR Green Flowcell Assay:
- Better QC method for higher density flowcells than 20X assay
- More indicative of cluster number on sequencer than 20X assay
- Reduction of fail rate on sequencers caused by poor quality flowcells
- Reduction of rework loop time

Linearization and Blocking QC Assay

Inefficient linearization and blocking of samples on a flowcell is a common fail mode in the Illumina cluster generation process. Failures in this part of the process caused by numerous factors, such as inadequate volume used, have lead to multiple fail modes (including red laser "washout/fog" and inefficient primer annealing) on the sequencer. To address this, a QC assay has been designed using a fluorescently labeled primer.



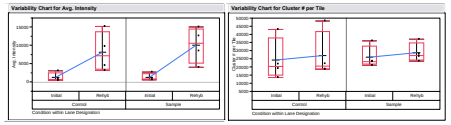
Typhoon Scan of Flowcells with P5-Cy3 Probe for Linearization and Blocking QC: A fluorescent primer is added using an abbreviated version of the Illumina Multi Primer Hybridization protocol then subsequently scanned on a Typhoon Scanner set for Cy3 fluorescence and analyzed using ImageQuant computer software. If there has been inefficient blocking, then the probe will bind to the unblocked complimentary P5 adaptor grafted on the flowcell and fluoresce.

This QC can currently be applied post-Linearization and Blocking or post-sequencing, especially if the run has been shown to exhibit significant background noise (i.e. – red laser washout/fog.)

Benefits Linearization and Blocking QC Assay:
- Shortens the rework loop for failed flowcells
- Identifies sequencing fail modes attributed to linearization and blocking from those attributed to the sequencing process itself
- Provides troubleshooting tool for problem flowcells
- Allows for experimental optimization of linearization and blocking chemistry

Flowcell Rescue on the Sequencer

Some flowcells exhibit unusually low first base intensity (average A intensity < 250 with 200 ms exposure). One cause of this could be poor primer annealing, due to low consumption or uneven flow of primer hybridization mix or 0.1 N NaOH on the cluster station or by the flowcell sitting too long in scan mix. A method to rescue these flowcells by re-primer annealing on the sequencer has been implemented.



Results of Flowcell Rescue using Primer Melt and Rehybridization on the Sequencer: At left, average A intensity before and after primer rehybridization. At right, cluster number found by the sequencer before and after primer rehybridization. The rescue is performed using same script as primer annealing for Paired End reads (with extra line attached to Port 8 on Vici valve), and melts off original primer then reanneals using new primer mix.

Benefits of Primer Melt and Rehybridization:
- Reduces overall fail rate by rescuing otherwise failing flowcells
- Reduces rework by rescuing flowcells at the sequencer stage, avoiding the need to re-amplify the samples on new flowcells
- Can increase yield by increasing the number of clusters found by the sequencer
- Potentially could be used to reduce load on cluster stations

Internal Controls

In the current flowcell setup, 1 lane is dedicated to a control sample (PhiX Control Library), which consume 12.5% of read capacity. A set of internal controls to be added to every lane of every flowcell will eliminate the need for the control lane.

(IC1) CACGTCACGTGTACGTAGTACGTACGCATACACGCATGTACGTATATAGTACTCGCGTAG
(IC2) TCTAGAGTACAGTGACGAGTGACGATACACACGCGTCTCTGACGCGCGCGATAGTATCATC
(IC3) GTGCAGTGCGCACATATCACATGCTATGCGTGTATGACAGTCGTATACTGCGAGTGAGCA
(IC4) AGATCTCATATCGTCGCTCGTCATGCGTGTGTATACGAGCATACGCGTACGTCACATCGT

The internal control sequences are 4 cloned monotemplates with balanced base composition and not PCR amplified. Each base occurs exactly once at a given sequencing cycle, and they are spiked in at 100k per lane (~1% of clusters).

Benefits of Internal Controls:
- Increases read capacity by 12.5% by eliminating control lane
- Allows collection of reliable metrics and improves data quality
- Provides sequencing run performance metrics independent of library quality
- Allows sequencing quality and accuracy comparisons lane by lane and tile by tile

Summary

In conclusion, we have performed extensive failure mode analysis and process mapping of our Illumina Genome Analyzer line. Based on the results of these investigations, we have developed a variety of quality control assays and process improvements that provide a more consistent data output and improved visibility over the entire process. This has allowed scaling up of the line over the past year to proceed more smoothly.

As a result of these quality improvements and implemntation of QC checkpoints, we have been able to greatly increase the quality of the process while increasing the throughput:

	January 2007	January 2008
Number of Instruments	8	20
Number of Flowcells per Day	Up to 2	Up to 8
Average Number of Runs per Week	5	25+
Run Fail Rate	>50%	<10%