

Reports

Lane-by-lane sequencing using Illumina's Genome Analyzer II

Michael T. Gravina, Jenny H. Lin, and Stuart S. Levine

Departments of Biology and Biological Engineering, Koch Institute for Integrative Cancer Research and MIT Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

BioTechniques 54:265-269 (May 2013) doi 10.2144/000114032

Keywords: high-throughput DNA sequencing; Illumina genome analyzer; lane-by-lane sequencing

Supplementary material for this article is available at www.BioTechniques.com/article114032.

Next-generation sequencing has become an essential tool in molecular biology that has been successfully applied to a broad variety of experimental approaches. While several platforms for next-generation sequencing exist, the most commonly used approach is sequencing-by-synthesis, implemented on Illumina's Genome Analyzer II (GAII) and HiSeq2000 systems. A key constraint of these sequencers is the need to run multiple lanes of samples with identical parameters as part of a single flowcell. Here, we present a series of modifications to the Illumina Genome Analyzer II, along with a script generating tool, that allow users to run the GAII in a lane-by-lane manner. Any number of lanes can be run at one time. Repeated use of the same flowcell on multiple sequencing runs does not appreciably reduce the intensity, cluster density, or accuracy of the run. These modifications will enable smaller-scale experiments with unusual design parameters to be run routinely on the GAII.

Since the introduction of the technology nearly a decade ago, high-throughput sequencing has dramatically changed the approaches used to study biological systems (1). These include global and targeted re-sequencing for SNP (2) and CNV characterization (3), ChIP-seq (4), transcriptome sequencing (5), and metagenome studies (6). Illumina's platforms, including the Genome Analyzer II (GAII) and the HiSeq2000, use a sequencing-by-synthesis approach in which a single nucleotide is added to very large numbers of unique templates on a flowcell and the entire flowcell is imaged before adding a new nucleotide. This massive parallelization of sequencing has allowed for dramatic reductions in the cost of sequencing. Additional cost-efficiency, rapid sample turnover, and proper controlling of complex samples have been achieved through the use of parallel sequencing channels prepared and sequenced simultaneously on each optical flowcell.

A critical limitation of the Illumina system is the need to run eight lanes simultaneously and with identical parameters (read-length, indexing, and chemistry type). This limitation is especially challenging on the older GAII machines that provide fewer bases per dollar than the HiSeq2000, often making it difficult to find enough samples to "fill out" a flowcell. Here we present a series of simple modifications to the Illumina GAII instrument and XML scripts that allow the sequencer to run in a lane-by-lane manner. Flowcells can be clustered on any desired number of lanes using the Illumina cBot, sequenced, stored, and then clustered again on remaining lanes without appreciable reduction in intensity, cluster density, or accuracy of the run. Read quality is comparable with standard 8-lane runs after paired-end turnaround chemistry to a length of 80 nucleotides.

Materials and methods

Clustering

Lane-by-lane clustering is performed on the Illumina cBot. Clustering kits (Part# GD-300-2001 and PE-300-2001 for single read and paired-end, respectively, Illumina, Inc., San Diego, CA) were modified by labeling reagent tubes with their reagent number. Tubes corresponding to unused lanes were cut away while frozen (Figure 1a). Unused reagents can be stored frozen for future runs. Empty wells on the reagent plate were replaced with tube strips (Part# MP73004L, Nova Biostorage Plus, Canonsburg, PA) filled with high salt buffer (Solution PR1 in all SBS kits for the Genome Analyzer, Illumina, Inc.) to prevent dehydration of the flowcell. Lane 1 was always included in the first set of lanes clustered on a given flowcell to permit edge-finding and tilt-setting.

Method summary:

We have modified Illumina's GAII to run partial flowcells through minor changes in the instrument and fluidics scripts. These changes are suited for rapid protocol prototyping and sequencing experiments with unusual chemistry requirements.

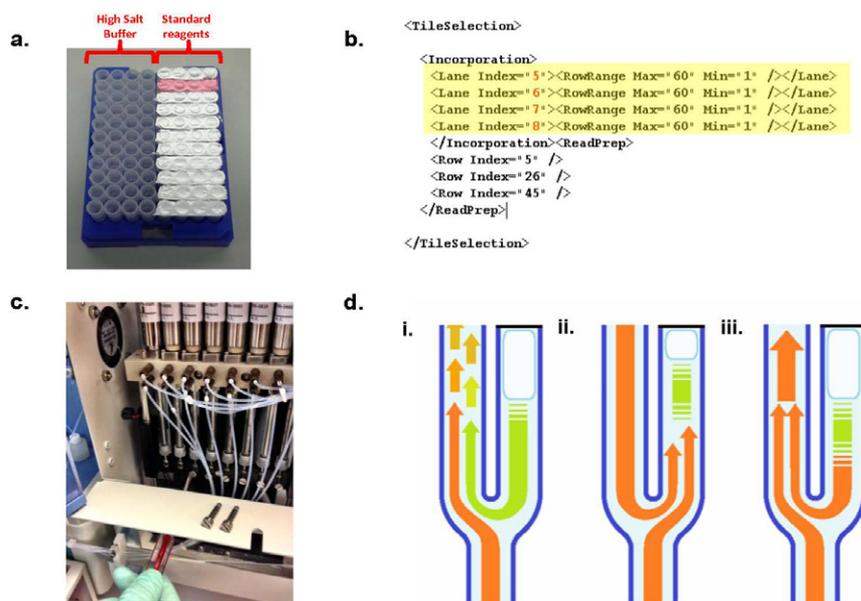


Figure 1. Setting up lane-by-lane sequencing on an Illumina GAIIx. (a) cBOT cluster plate set up to cluster lanes 5–8. Lanes 1–4 have been removed using a razor blade and replaced with vials containing high salt buffer. (b) The TileSelection section of an XML recipe file for 4-lane sequencing of the flowcell clustered in (a). Only lanes 5–8 will be imaged. (c) Disconnecting the syringe-pumps. Once the manifold housing is removed, the plunger heads can be unscrewed from the motorized crossbar. Here, pumping has been inactivated for lanes 1 and 2. (d) Pause-and-pulse reagent delivery method for reads longer than 40 nucleotides. The left lane is active and the right lane is disconnected. Orange arrows represent flow of new reagent, green arrows represent movement of previous reagent, the white ovals represent a small air bubble remaining in the disconnected lane. i. During the initial pumping stage, reagent delivered from the inlet (bottom) may be contaminated by fluid from disconnected lanes as small bubbles expand due to negative pressure from the pump. ii. After pumping, a 5 s pause allows a small amount of the current reagent to back-fill through the inlet manifold as pressure equalizes in the system. iii. A 20 μ L pulse then finishes delivery with homogenous reagent.

Imaging

Imaging was limited to desired lanes in order to reduce run time. The XML recipe files were edited to remove the “Lane Index” line entry for each unused lane under the “TileSelection” header (Figure 1b). Note that while this does not prevent the brief imaging during calibration, full within-read imaging is restricted to the active lanes.

Reagent delivery

In order to preserve high signal-to-noise ratios during longer runs, salt solution in the inactive lanes was removed. Prior to each read, deionized water was flushed across the entire flowcell while the reagent manifold was still in the standard configuration, pre-charging the inactive lanes with salt-free medium.

Following the water wash, reagent flow was physically limited to the desired lanes by inactivating the pumps corresponding to the unused lanes. The plunger ends of the syringe-pumps for the unused lanes were disconnected from the motorized crossbar by removing the connecting screws while in the top position (Figure 1c).

The reagent delivery volume settings were changed to compensate for the effect of fewer pumps on rates of flow.

The default delivery volumes in the XML recipe file (the “PumpToFlowcell” variables of each “Chemistry Name” item under the Chemistry Definitions) were transformed using the following formula:

$$V_C = V_0 + V_1 \left(\frac{8}{L} - 1 \right)$$

where V_C is the nominal volume after compensating for fewer pumps, V_0 is the default volume listed in the XML file, V_1 is the inlet line volume for the given reagent, and L is the number of active lanes.

In cases where a chemistry definition calls for the same solution multiple times in a row, only the first volume is transformed. In addition, the volumes for solutions 13, 14, and 15, which are delivered under the “Resynthesis” chemistry name item for paired-end recipes only, are not transformed using this formula but are instead multiplied by a factor of 1.5.

Inlet line volumes were determined experimentally for each reagent. For our GAII, they were approximately 150 μ L for reagents housed on the GAII (solutions 1–7) and 550 μ L for reagents on the paired-end module (PEM, solutions 9–27). These correspond to 15 and 55 units in the XML

recipe respectively. Line volumes may vary between instruments and therefore we recommend confirming them for each specific machine.

Several sources of cross-lane contamination were identified and mitigated during initial testing. The system was thoroughly washed prior to each run to eliminate any bubbles in the syringe-pumps that, by expanding and contracting during pump cycles, could draw fluid between active and inactive lanes through the inlet manifold (Figure 1d). On longer runs (>40 nucleotides per read), 400 μ L of additional volume was added on top of the compensated volumes in order to facilitate in-lane mixing, followed by a 5 s pause after each delivery to equilibrate the current solution into the inactive lanes, followed by another 200 μ L pulse to complete the delivery with non-contaminated reagent. We refer to these additional steps for long read quality as “pause-and-pulse” delivery. When used, these changes do not apply to PEM reagents.

Paired-end reagents

Paired-end reagents are thawed and divided into aliquots for multiple uses. Reagents that are not used for the current run can be divided into separate tubes and flash-frozen in liquid nitrogen one time without negatively affecting their performance in paired-end turnaround chemistry.

Recipe generation

We developed a Python script (for Python 2.7 or later) that converts existing recipes into equivalent versions running an arbitrary selection of lanes in order to avoid human error and facilitate generation of the large variety of possible lane-by-lane recipe files. Lane imaging and reagent delivery volumes are automatically adjusted, run times are estimated, and user alerts are inserted to remind operators to disconnect and reconnect syringe pumps at appropriate times. Pause-and-pulse delivery for long reads is optional. For easy identification, the output XML files are named based on their parameter selection.

By default our Python script implements a switch to two-lane operation during the turnaround phase of any one-lane run, sacrificing a non-sequencing lane to conserve time and reagents as we found that paired-end resynthesis chemistry with only one active lane had prohibitively high consumption. This option can be turned off at recipe generation.

Results and discussion

Initial experiments with food dye were used to validate lane-by-lane control of reagent

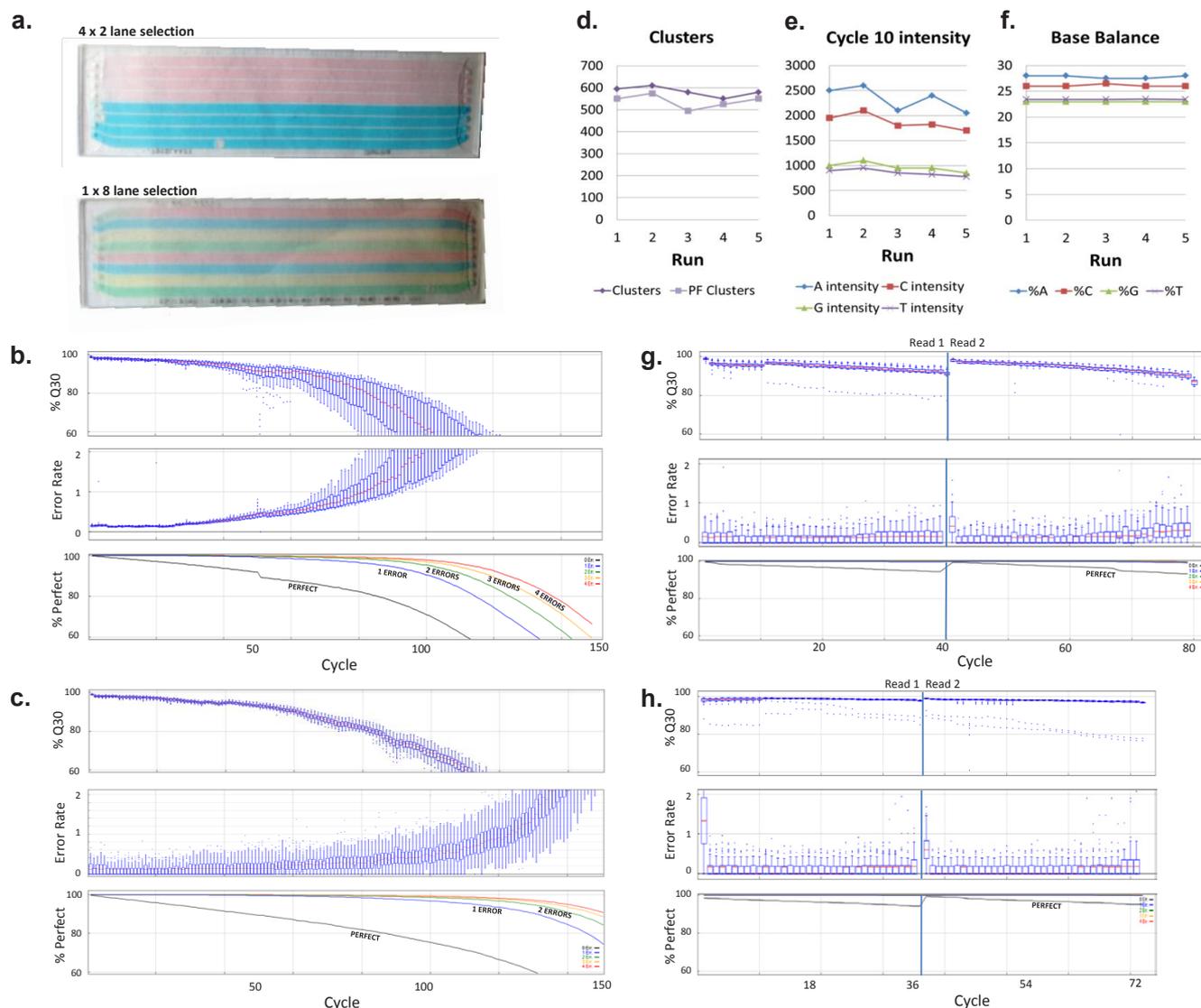


Figure 2. Lane-by-lane sequencing produces high quality sequence data. (a) Demonstration of lane-by-lane reagent delivery. Water marked with food coloring was delivered to either half of an Illumina GAII flowcell (top) or single lanes (bottom). No cross-lane mixing is observed. (b) Quality control data from the phiX lane of a 150 nucleotide single-end run performed on 2 lanes. Tile based quantiles for Q30 scores (top), quantiles for tile based phiX error rates (middle) and the cumulative number of accumulated errors in each read (bottom) are shown for each cycle. (c) Quality control data from the phiX lane of a 150 nucleotide single-end run performed on 8-lanes. Data shown as in (b). (d-f) Flowcell reuse does not negatively impact performance. The ability to reuse flowcells was tested by repeated clustering and sequencing of single lanes over the course of two weeks. Number of detected clusters and clusters passing Illumina quality metrics (d), intensity of each base at cycle 10 (e), and base frequencies at cycle 10 (f). (g) Quality control data from the phiX lane of a 40/40 paired-end run performed on two lanes. Data shown as in (b). (h) Quality control data from the phiX lane of a 36/36 paired-end run performed on 8 lanes. Data shown as in (b).

distribution. Single or multiple syringes were left connected to the motorized crossbar and diluted food coloring was injected across the flowcell. No mixing or color bleed was observed in preliminary experiments suggesting that disconnecting the syringes was a viable mechanism for lane-by-lane control (Figure 2a).

We next validated the method for lane-by-lane sequencing on a single control sample (Figure 2b). Control Illumina libraries derived from the bacteriophage phiX were clustered on 2 lanes of a standard GAII single end flowcell and sequenced for 150 nucleotides (~3 days) using pause-and-pulse delivery. Error rate and quality plots

show strong initial quality of the read with the median error rate remaining well below 1% out to 80 nucleotides and below 2% through 100 nucleotides. Above 100 nucleotides, the error rate climbed rapidly. The increase in error rate occurs more rapidly than that seen during standard full flowcell GAII runs (Figure 2c).

We next evaluated the effects of repeated use of a flowcell on basic sequencing metrics. In order to simulate the partial re-use conditions that would occur normally in the laboratory, single lanes of a Genome Analyzer single-read flowcell v2 were alternately clustered with phiX and sequenced over the course of

two weeks. Lanes 1–4 were clustered and run sequentially followed by a 4-lane run using the rest of the flowcell. Sequencing reagents were changed after the third run and pause-and-pulse delivery was not used due to the shortness of the read.

Data quality remained similar over the course of the five cluster-sequence cycles. Neither absolute cluster density nor density of clusters passing Illumina's informatic filters showed a decrease over time. Both cluster density metrics varied less than 20% across all clusters, suggesting that repeated cycles of washes on the flowcell do not negatively impact the clustering process (Figure 2d).

| Number of lanes | Pause-and-Pulse | nt per kit* | turnarounds per kit** |
|-----------------|-----------------|-------------|-----------------------|
| 1 | no | 140 | 2 |
| 2 | no | 106 | 2 |
| 3 | no | 82 | 2 |
| 4 | no | 66 | 2 |
| 5 | no | 55 | 1 |
| 6 | no | 47 | 1 |
| 7 | no | | N/A*** |
| 8 | no | | N/A*** |
| 1 | yes | 81 | 2 |
| 2 | yes | 54 | 2 |
| 3 | yes | 41 | 2 |
| 4 | yes | 33 | 2 |
| Five or more | yes | | N/A*** |

*Per 1 40 nucleotide SBS kit
 **Per 1 paired-end cluster kit
 ***Less costly to run a standard 8-lane flowcell

Repeated use of the same flowcell over multiple sequencing runs did not appreciably reduce per-base intensities, which diminished less than 20% for A and C and less than 10% for G and T (Figure 2e) and base frequencies varied less than 5%, as expected (Figure 2f).

Finally, we evaluated the performance of partial flowcells in paired-end sequencing by clustering genomic DNA on a 4-lane, 40 nucleotide by 40 nucleotide paired-end flowcell using pause-and-pulse delivery and turnaround reagents that had been divided and flash-refrozen once (Figure 2g). The median error rate remained below 1.0. The percentage of perfect reads remained above 80% and the median Q-score remained above 35. These metrics apply to both forward and reverse reads and differ less than 20% from the same metrics for an 8-lane, 35 nucleotide by 35 nucleotide paired-end comparison flowcell (Figure 2h).

The modifications outlined above enable GAI users to routinely perform small-scale experiments that were previously constrained by the grouped operation of the flowcell lanes. This method can be used for an arbitrarily small number of lanes, eliminating the need to wait for eight lanes of matching parameters before clustering and increasing the per-base speed of sequencing as the number of lanes is decreased (Table 1). Flowcells have been stored for up to one month in a partially clustered state without observable decreases in the performance of the unclustered lanes (data not shown).

Lane-by-lane sequencing allows for high quality metrics for up to 80 nucleotides on each read and is readily adapted for multi-lane experiments requiring different chemistries for each lane. This

flexibility allows the GAI to efficiently handle unusual experimental designs. Applications for which it has been successfully utilized in our lab include lane-by-lane priming of the reverse read of a paired-end run, which is impossible in the native operation of the instrument due to the universal delivery of a single read two primer mix to all lanes. The HiTS-FLIP method is also highly amenable to lane-by-lane sequencing, since each lane can now receive a separate flow of protein for binding affinity measurements, independent of the other lanes (7). While the method is limited to reads of 80 nucleotides, further refinement may enable longer read lengths. We also note that the Illumina HiSeq uses a similar fluidics architecture to the Genome Analyzer. As such, the physical modifications we present here should be readily adaptable to the higher throughput system and lane-by-lane sequencing may be achievable on the HiSeq.

Acknowledgments

We are grateful to members of the MIT BioMicro Center and Sarah Adelman and Thom Theara (Illumina, Inc.) for helpful discussions and comments on the manuscript. This work was funded by the National Cancer Institute of the NIH under award P30-CA14051 and the National Institute of Environmental Health Sciences of the NIH under award P30-ES002109 as well as the Departments of Biology and Biological Engineering at MIT (S.S.L.).

The authors wish to dedicate this paper to the memory of Officer Sean Collier, for his caring service to the MIT community and for his sacrifice.

Competing interests

The authors declare no competing interests.

References

- Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380.
- Nielsen, R., J.S. Paul, A. Albrechtsen, and Y.S. Song. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12:443-451.
- Ku, C.S., E.Y. Loy, A. Salim, Y. Pawitan, and K.S. Chia. 2010. The discovery of human genetic variations and their use as disease markers: past, present and future. *J. Hum. Genet.* 55:403-415.
- Furey, T.S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* 13:840-852.
- Ozsolak, F. and P.M. Milos. 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12:87-98.
- Shokralla, S., J.L. Spall, J.F. Gibson, and M. Hajibabaei. 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* 21:1794-1805.
- Nutiu, R., R.C. Friedman, S. Luo, I. Khrebtukova, D. Silva, R. Li, L. Zhang, G.P. Schroth, and C.B. Burge. 2011. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat. Biotechnol.* 29:659-664.

Received 10 March 2013; accepted 17 April 2013

Address correspondence to Stuart Levine, BioMicro Center, Massachusetts Institute of Technology, Cambridge. E-mail: slevine@mit.edu

To purchase reprints of this article, contact: biotechniques@fosterprinting.com