

# Reversing the Trend: More Bases with Less Computing

Software advancements reduce computing requirements while accommodating the rapidly expanding sequencing output with the Genome Analyzer<sub>IIx</sub>.

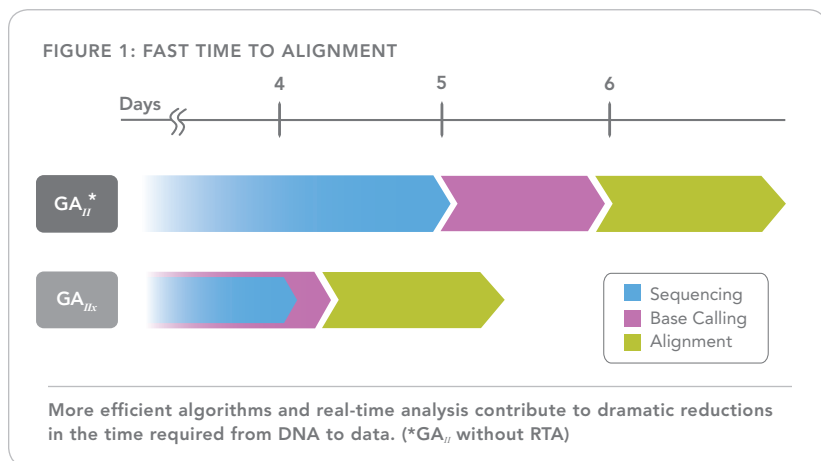
## INTRODUCTION

The output of Illumina sequencing technologies continues to expand rapidly. Incorporating changes in hardware, chemistry, and software, the Genome Analyzer<sub>IIx</sub> represents the latest milestone in sequencing output. Enormous gains in data output trigger the need for streamlined analysis methods and require solutions that make efficient use of computer hardware. With Illumina's recent software advancements, researchers can expand the scope of their sequencing projects.

Illumina is leading the industry by driving down the computational burden of processing data in the face of rapidly increasing sequencing output. Yields have scaled 15-fold over the course of 2008, and are on a path to reach 95 Gb by the end of 2009. If computational infrastructure requirements grew proportionately, large genome-scale experiments would remain out of reach to most researchers. However, the Genome Analyzer<sub>IIx</sub> with real-time primary analysis dramatically reduces the time, hardware, networking, and laboratory infrastructure requirements needed to produce mappable data.

## SHORTER PATH TO ALIGNED DATA

An important measure of a sequencing system's utility is its overall efficiency. For example, the amount of mappable data produced per day



and the total cost of installing and running the instrument are critical aspects of an economical sequencing solution. To minimize the time to generate mappable data and to increase the efficiency of the system, real-time base calling is now enabled on the standard Genome Analyzer instrument computer, and runs in parallel with sequencing chemistry and imaging (Figure 1). This allows much of the data processing and analysis to be completed before the sequencing chemistry finishes. Additionally, it can reduce the storage required for intermediate data and limit the amount of data that needs to travel across the network.

## REDUCED INFRASTRUCTURE

Reducing raw sequencing data into a compact base and quality format reduces the need for large data stor-

age hardware, reduces the load on the network, and minimizes the cost of IT support. More importantly, it reduces the time to aligned data so that more time can be devoted to biological analysis of the data (Figure 1).

## Reduced Network Utilization

While sequence output has increased, the data per run transferred from the Genome Analyzer to the network and to secondary analysis processing hardware has decreased by orders of magnitude (Figure 2). By summarizing data on the instrument computer, network loads are dramatically reduced. Without these on-instrument, off-network data reduction techniques, the image output of a fleet of Genome Analyzers would cripple most networks. For example, if a lab chose to export image files, it could run no more than five instru-

TABLE 1: DATA SUMMARIZATION EXPANDS LABORATORY SEQUENCING CAPACITY

OUTPUT TYPE	SINGLE GA <sub>IIx</sub> OUTPUT	SINGLE GA <sub>IIx</sub> OUTPUT	NETWORK BANDWIDTH	% NETWORK UTILIZATION	INSTRUMENT SATURATION POINT
Images	500 GByte/day	46 Mbit/s	1000 Mbit/s	50%	10
Bases and Qualities	15 GByte/day	1.4 Mbit/s	1000 Mbit/s	50%	360

ments concurrently without saturating a standard gigabit connection to a single file system (Table 1). Thanks to advances in real time data analysis, huge aggregate sequence output can now be supported with conventional network and hardware configurations.

#### Reduced Hardware Requirements

The widespread adoption of the Genome Analyzer has been driven in part by its ease to use, support for a range of applications, and suitability for virtually any lab environment. Illumina's highly efficient algorithms have allowed significant

analysis functionality to be added to the simple workstation that controls the Genome Analyzer. This reduction in the requirements for computational hardware has several practical benefits that will become even more important as sequencing output levels continue to increase. By performing image analysis and base calling on a simple tower, heat production, laboratory footprint, and power consumption are kept to a minimum. Other sequencing technologies have recently ramped up their computing infrastructure for primary analysis, with up to five times more process-

ing power, leading to commensurate increases in heat output and power consumption.

The computational efficiency of the Genome Analyzer enables customers to increase their sequencing throughput while keeping server hardware expenses to a minimum.

#### CONCLUSION

Because the volume of data that Illumina sequencing generates will continue to increase, Illumina is committed to providing scalable and manageable solutions for data management. Ultimately, the generation of more raw sequence will increase the need for secondary analysis and data storage, making optimization of data transport and storage critical. Illumina's unique efficiency and ease-of-use in processing sequencing data enable researchers to sequence more samples at greater depth than ever before.

#### ADDITIONAL INFORMATION

For more information about Illumina Sequencing, please visit [www.illumina.com/sequencing](http://www.illumina.com/sequencing) or contact us at the address below.

Illumina, Inc.

Customer Solutions

9885 Towne Centre Drive  
San Diego, CA 92121-1975

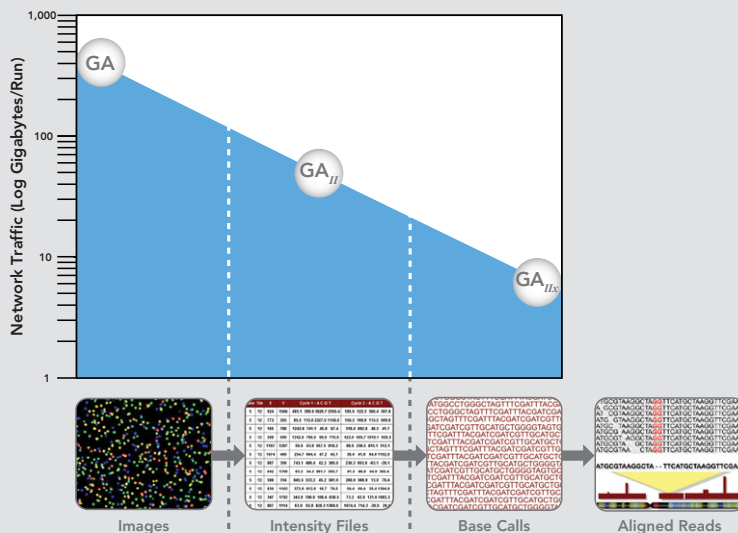
1.800.809.4566 (toll free)

1.858.202.4566 (outside North America)

[techsupport@illumina.com](mailto:techsupport@illumina.com)

[www.illumina.com](http://www.illumina.com)

FIGURE 2: DECREASING NETWORK BURDEN WITH SYSTEM ADVANCES



Real-time analysis generates base call-level data instead of bulky images or intensity files. This is performed on the instrument workstation, reducing the need for large file transfer over networks and complex data storage infrastructure.

#### FOR RESEARCH USE ONLY

© 2009 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPPro, and GenomeStudio are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

Pub. No. 770-2009-016 Current as of 5 May 2009

