

Illumina GA IIx & HiSeq 2000 Production Sequencing and QC Analysis Pipelines at the DOE Joint Genome Institute

Chris Daum¹, James Han¹, Matt Zane¹, Megan Kennedy¹, Matthew San Diego², Alex Copeland², Minkun Li², Susan Lucas¹

¹ Lawrence Livermore National Laboratory

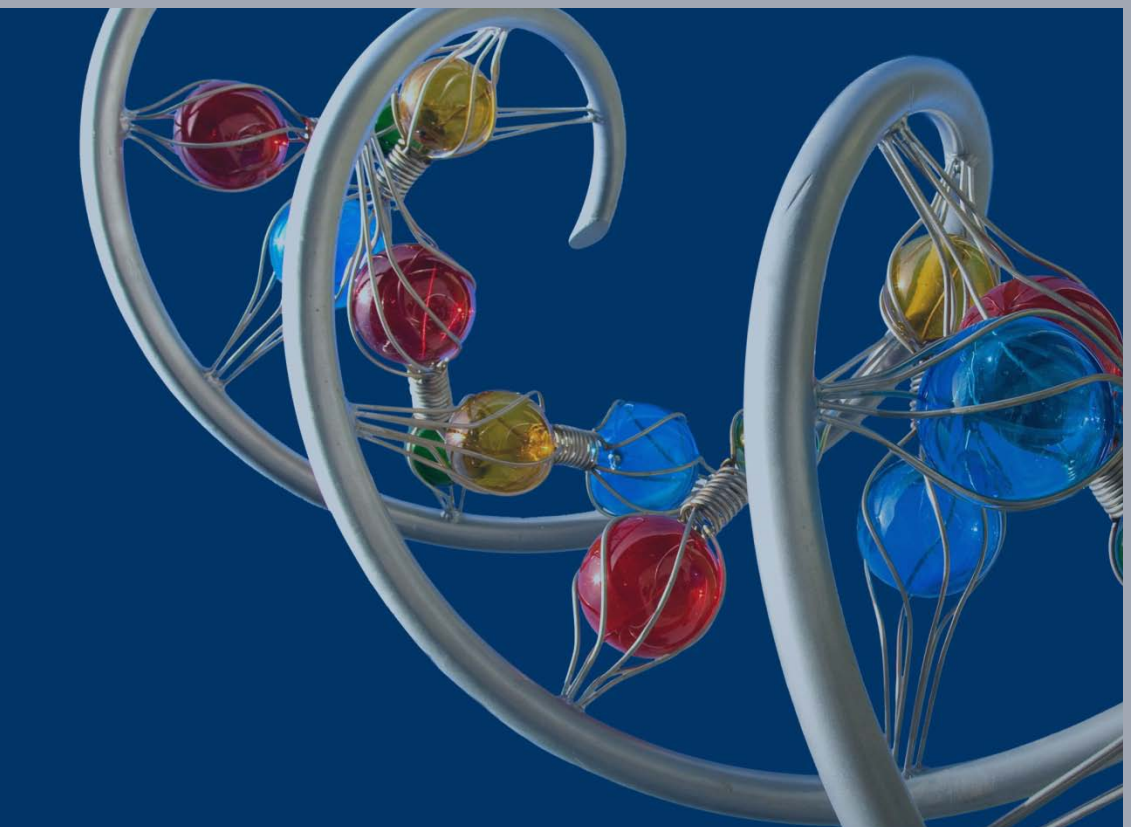
² Lawrence Berkeley National Laboratory

February 2011

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH112

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.



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 mission areas of renewable energy generation, global carbon management, and environmental characterization and clean-up. Within the JGI's Production Sequencing group, a robust Illumina Genome Analyzer and HiSeq pipeline has been established. Optimization of these sequencer pipelines has been ongoing with the aim of continual process improvement of the laboratory workflow, reducing operational costs and project cycle times to increase sample throughput, and improving the overall quality of the sequence generated. A sequence QC analysis pipeline has been implemented to automatically generate read and assembly level quality metrics. The foremost of these optimization projects, along with sequencing and operational strategies, throughput numbers, and sequencing quality results will be presented.

Introduction

The DOE Joint Genome Institute (JGI) was established in 1997 to unite the expertise and resources in genome mapping, DNA sequencing, technology development, and information sciences pioneered at the DOE genome centers of Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (LANL). In January 1999, high-throughput DNA sequencing began at the Production Genomics Facility (PGF) in Walnut Creek, CA.

Illumina Production Pipeline – Updates:

May 2010: Installation and utilization of SCS 2.8/OLB 2.8 software to adopt 95G GA/Ix runs.

August 2010: Two HiSeqs installed. This significantly increased Illumina Production throughput capacity.

December 2010: Three additional HiSeqs installed and three GA/Ix analyzers were decommissioned.

Upcoming: In February 2011, three additional HiSeq sequencers to be installed for a total of 8 HiSeq & 9 GA/Ix instruments.

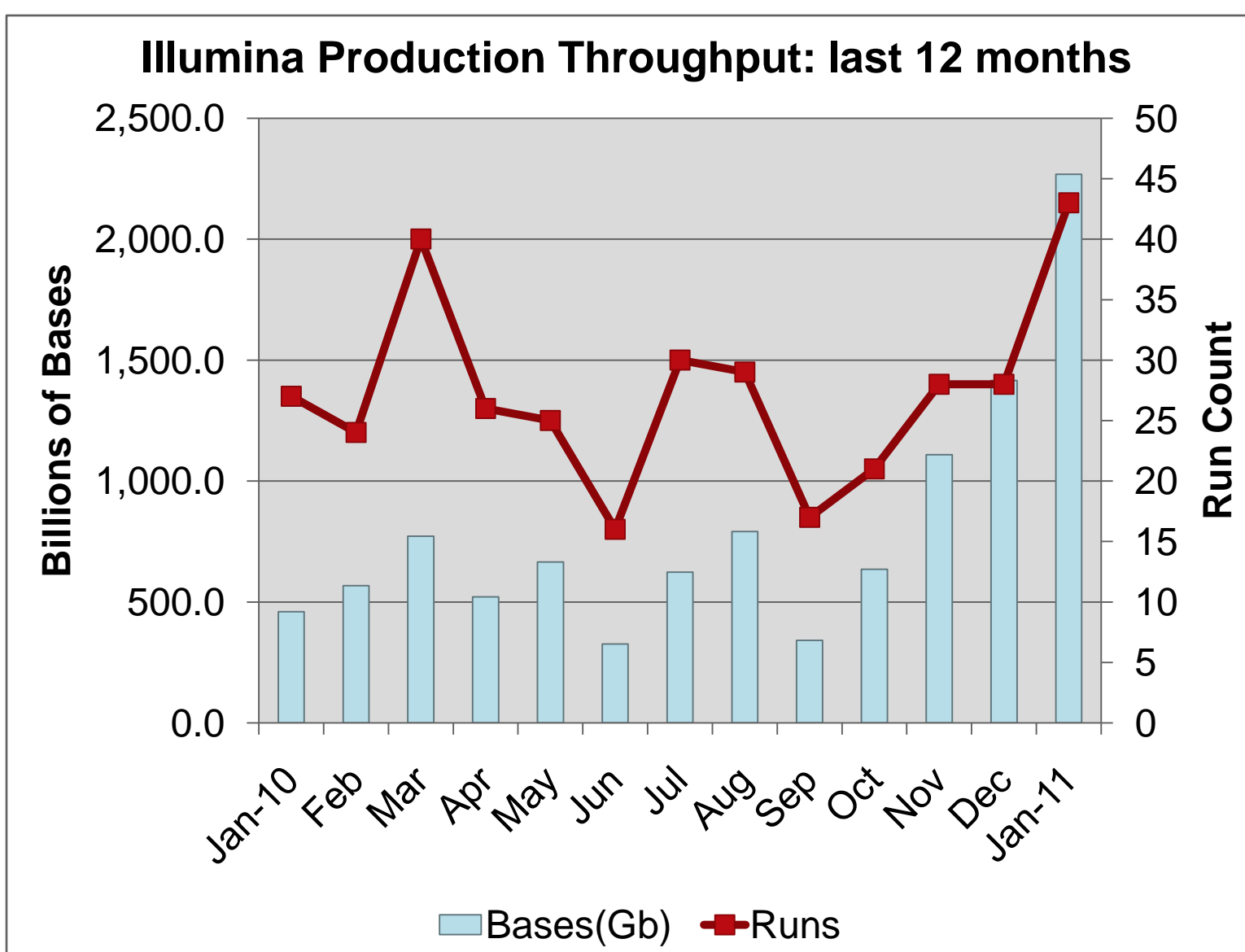


Chart 1: Illumina throughput of number bases collected & number of sequencing runs since Jan 2010

	Runs	Tb
FY10	274	5.63
FY11 YTD	123	5.43

Table 1: Fiscal Year 2010 & Fiscal Year 2011 YTD Production Throughput

	cBot	HiSeq	GA/Ix	PEM	Staff
FY10	6	2	12	12	5
FY11	6	8	9	9	7

Table 2: Fiscal Year 2010 & Fiscal Year 2011 instruments & staffing

Ergonomics: Program and Philosophy

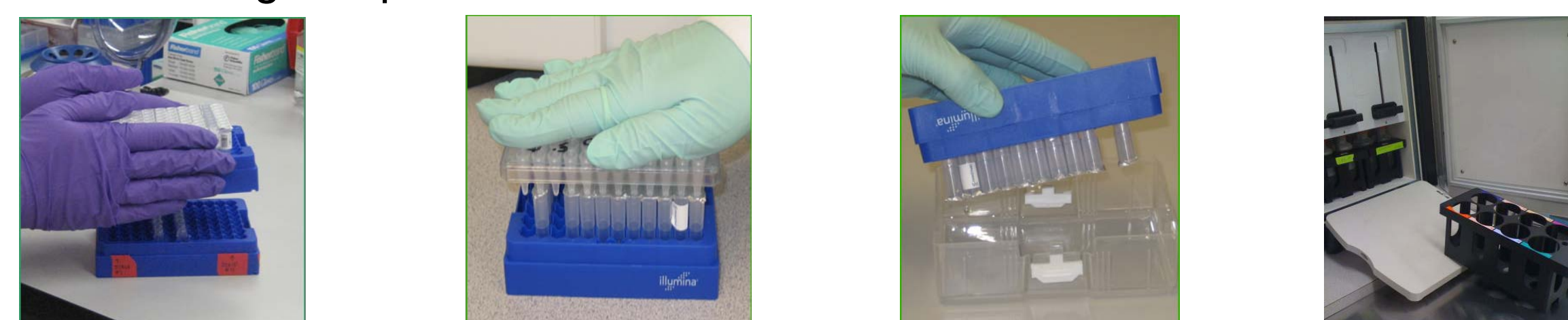
Employees at the JGI, whether working in an office, laboratory, or both are faced with repetitive and detail-oriented tasks daily. The staff, with management support, continuously seek to develop ergonomic and safety awareness, injury prevention and an education program to reduce ergonomic risk and repetitive strain injuries.

JGI's innovative ergonomic tools and educational programs were awarded top honors at the Institute of Industrial Engineers (IIE) Applied Ergonomics Conference. In 2010 the grassroots initiative Empowering Employees in Ergonomics to improve training, education and communication was awarded the Ergo Cup. Employees created a Safety Culture Working Group and Ergonomics Working Group as part of the program, which led to a 92% reduction in ergonomic-related incidents and 100% reduction in lost work days.



Illumina Ergonomics & Process Improvements

Continual improvement and innovation may happen in small steps that together have a large impact on ergonomics. Several tools were built by operators using recycled material, and others are low cost (less than 100 dollars) solutions. In March these economical solutions are competing for the 2011 IIE Ergo Cup.



Recycled plate used to remove strip tubes without a pinch grip.

96 well plate to easily pierce foil seals.

Kit lid used to collect excess reagents for disposal.

Monitor stand to slide in reagent cassettes.

JGI's education program is being upgraded to include instructional videos of all best practices.

Continuous Improvement Lean Six Sigma Program



Lean Six Sigma is a data driven process to improve production by reducing variation and eliminating waste. All JGI Production Illumina operators have completed Lean Six Sigma Greenbelt Certification. Currently they are working on the following Projects:

- **Improving Data Collection and Accessibility** - Working to reduce inconsistencies in data storage, and improve access to all collected data. This will allow real-time tracking of quality and run performance.
- **Tracking Form Consolidation** - This team worked to save time and streamline workflow by improving and reducing the amount of forms required in the daily Illumina process. This allowed for more efficient runs and reduced ergonomic risk due to excessive computer use.
- **Reducing Contamination and Variability** - Operators worked with Library Creation and Quality Control to reduce opportunities for contamination in the cluster generation process, and also worked to remove variability by implementing operator laboratory best practices.

Sequence Quality Assurance

The Quality Assurance program analyzes sequence prior to project assembly and scientific investigation. Rolling QC is an in-house sequence QC pipeline that performs a set collection of analyses and produces a summary report for each lane of Illumina data produced by the sequencing group. The pipeline calculates read quality, measures sequence uniqueness, and detects abnormal sequence motifs. An assembly, using Velvet, is used to measure coverage and detect contamination.

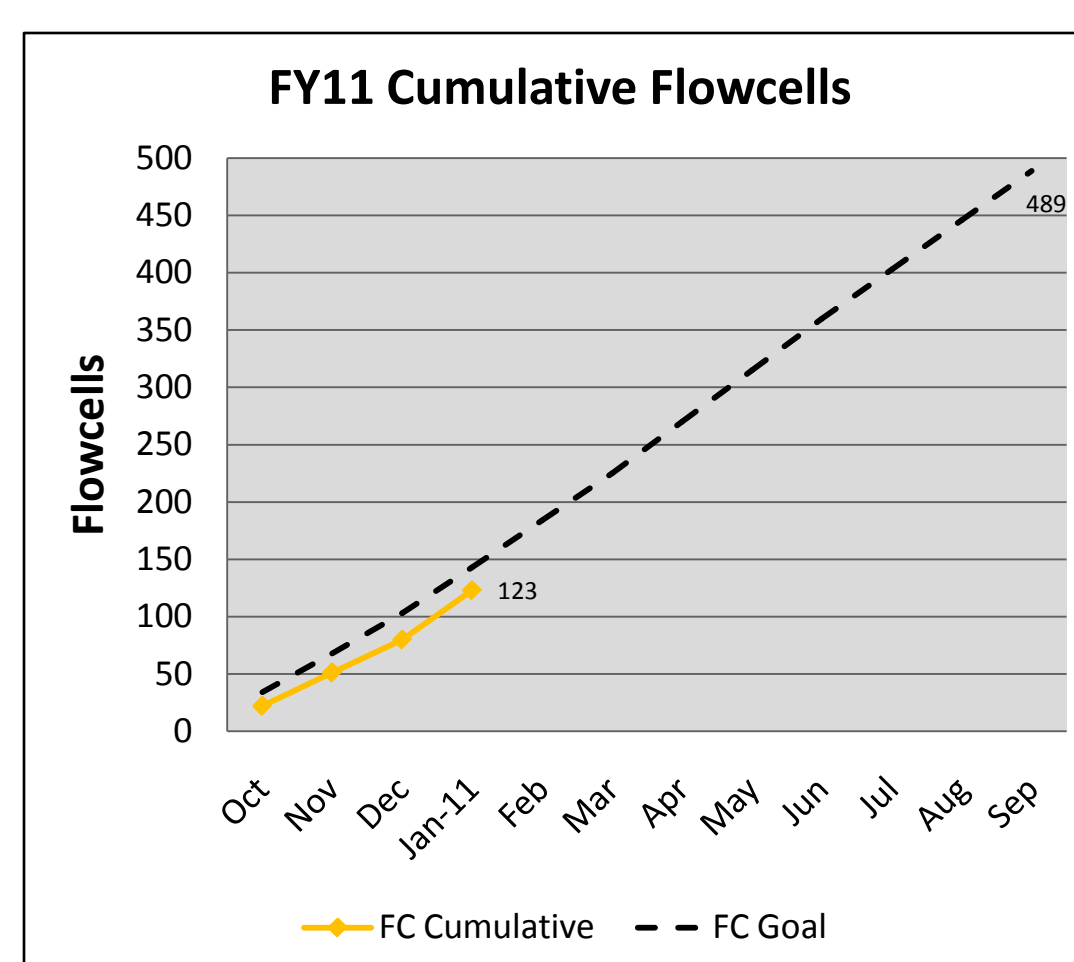
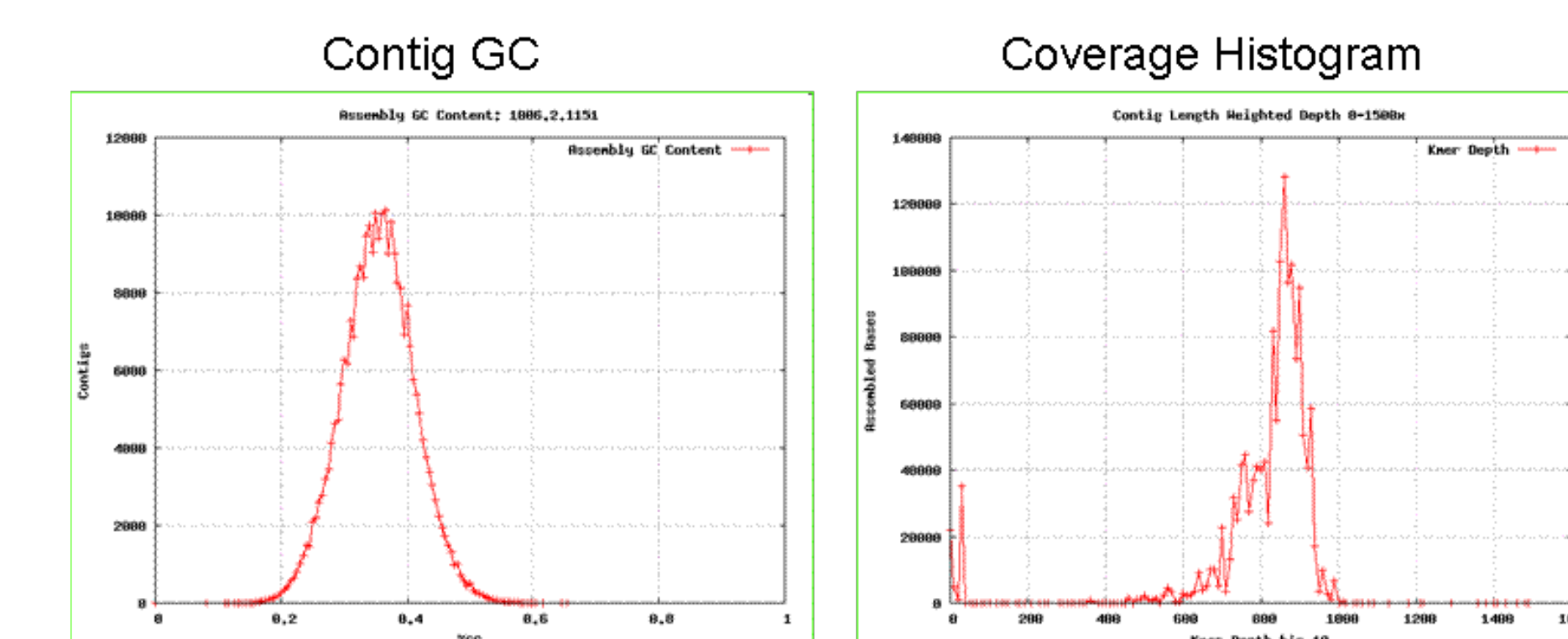
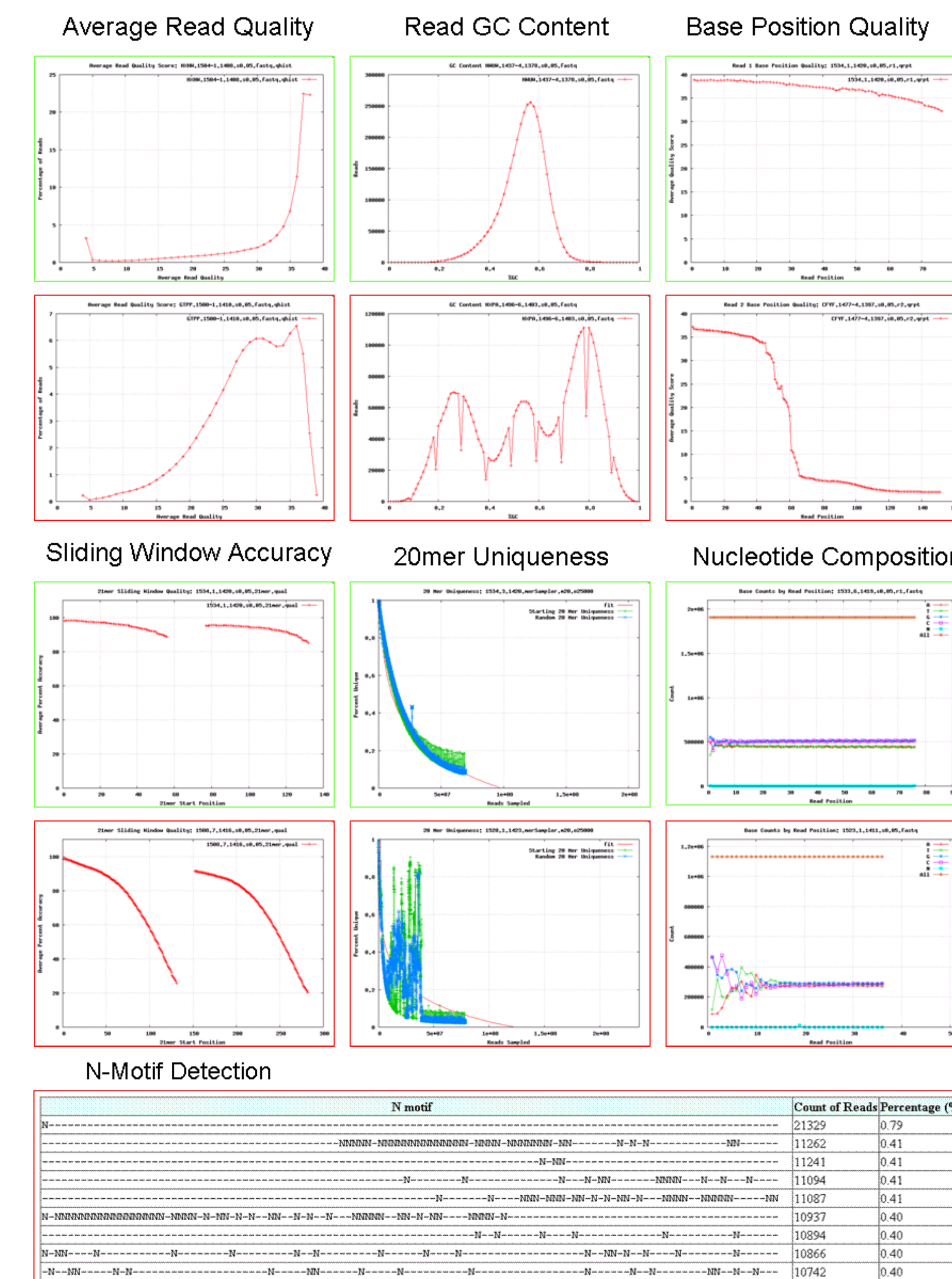


Chart 2 & 3: Fiscal Year 2011 (FY11) Throughput targets and cumulative actual to date

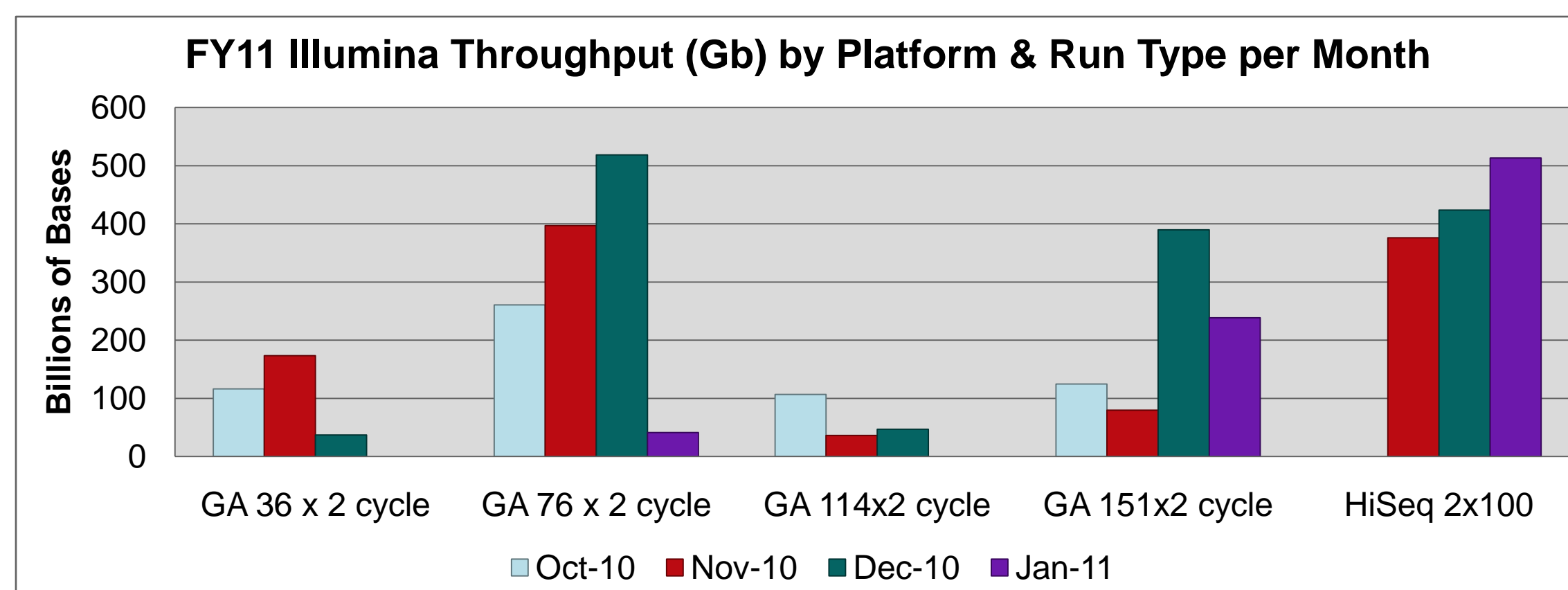
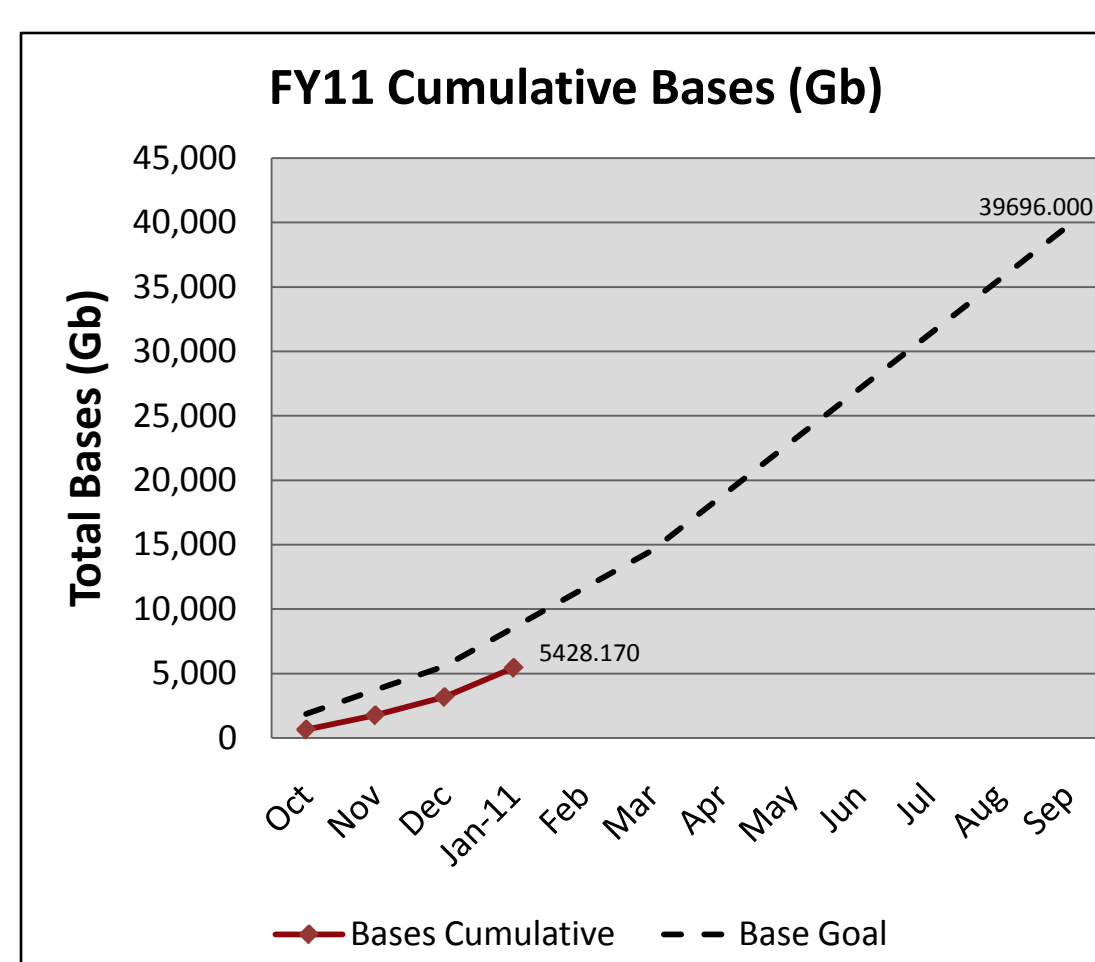


Chart 4: Fiscal Year 2011 (FY11) Total Bases (Gb) throughput by platform & run type