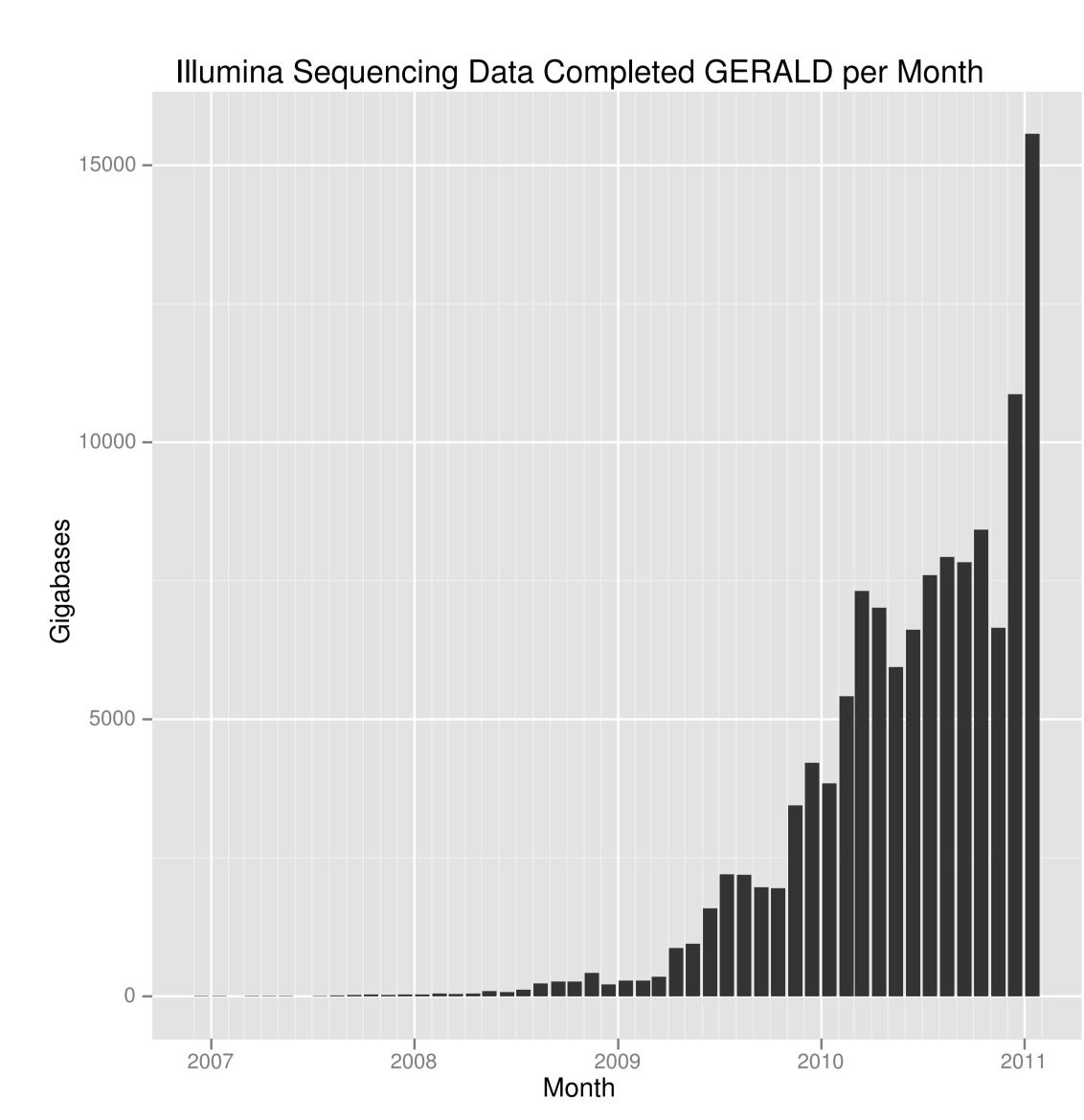
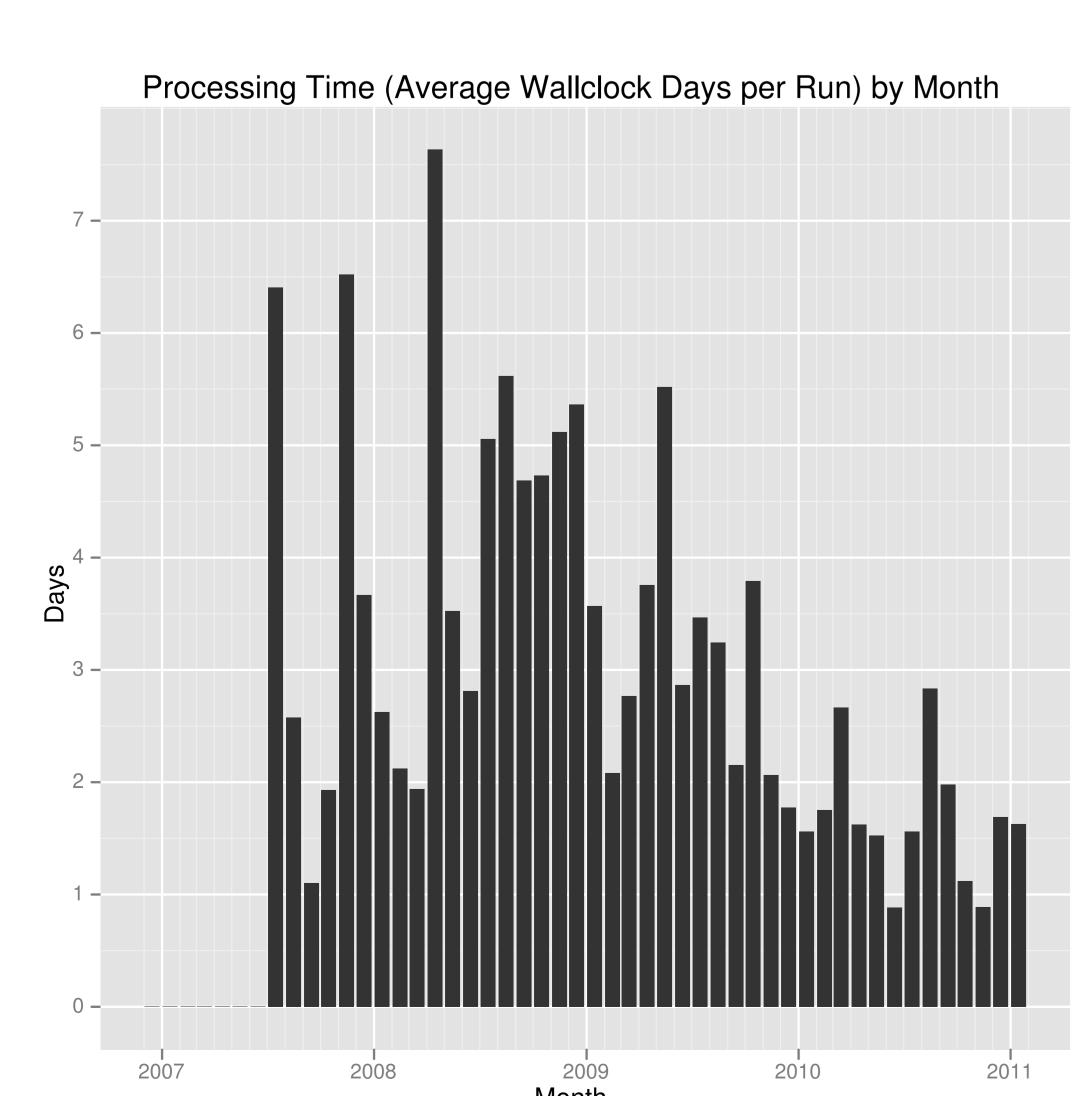
Large-Scale Illumina Sequencing and Analysis

Todd Hepler, William Schroeder, Scott Abbott, Michael Kiwala, Indraniel Das, Craig S. Pohl, David J. Dooling, Elaine R. Mardis, and Richard K. Wilson

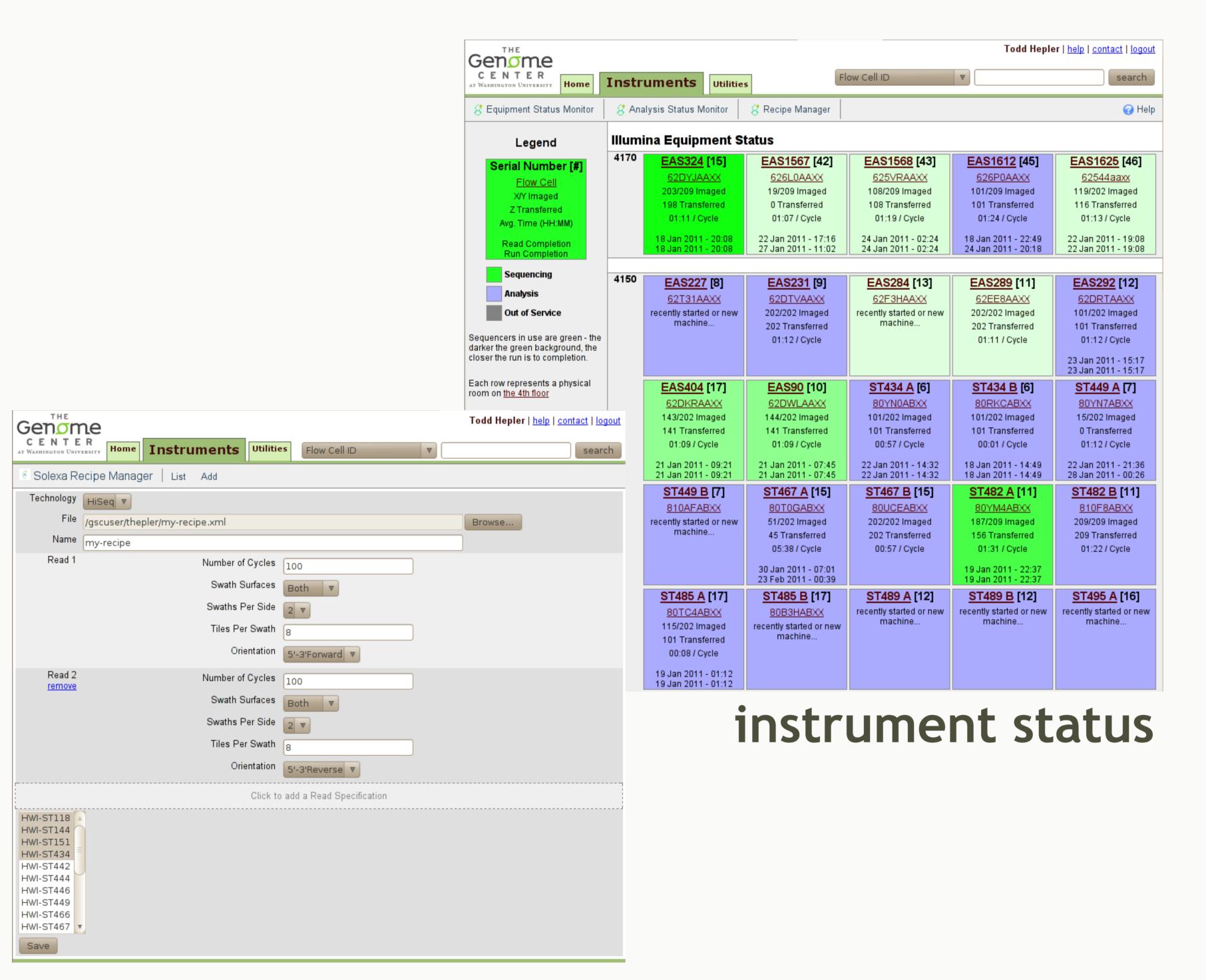
The Genome Center at Washington University in St. Louis

The Genome Center at Washington University has a largescale, high-throughput Illumina sequencing and primary analysis pipeline that has required the development of customized software to maximize throughput, increase quality control and minimize costs. In order to facilitate this system, we have developed a custom disk allocation system for managing data types and processing requirements, an automated interface to the SCS/HCS instrument software, a system for parallelizing the primary analysis and CASAVA secondary analysis by lane or sub lane (index), and a systematic way for generating quality control statistics. Analyzed run data and metrics are stored in a data-warehouse database and reported back to technicians and project managers through a number of custom views, including web status monitoring and automated reports. This analysis pipeline continues to evolve in response to challenges of scale due to new sequencing instruments, vendor code releases, system



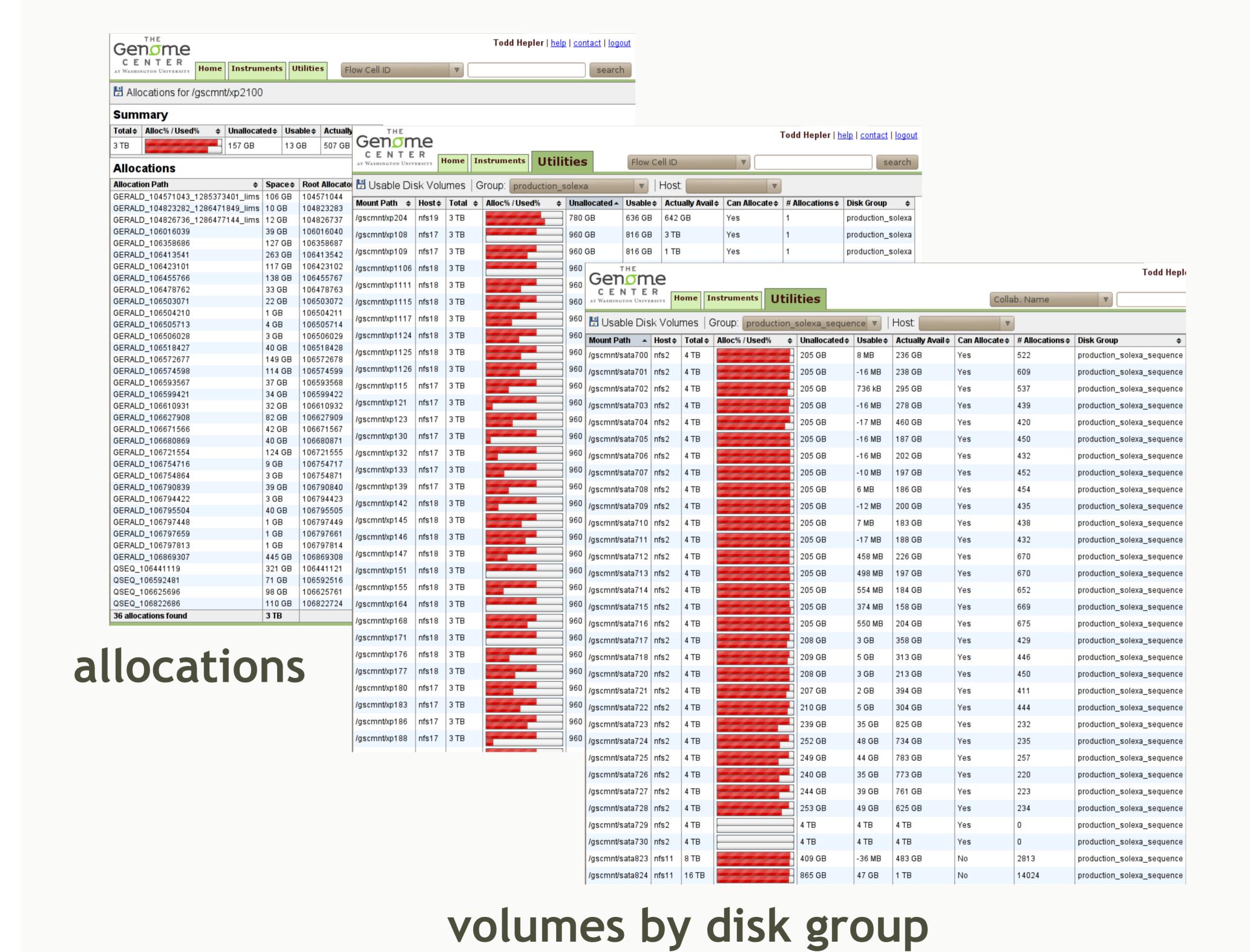


Instrument Management

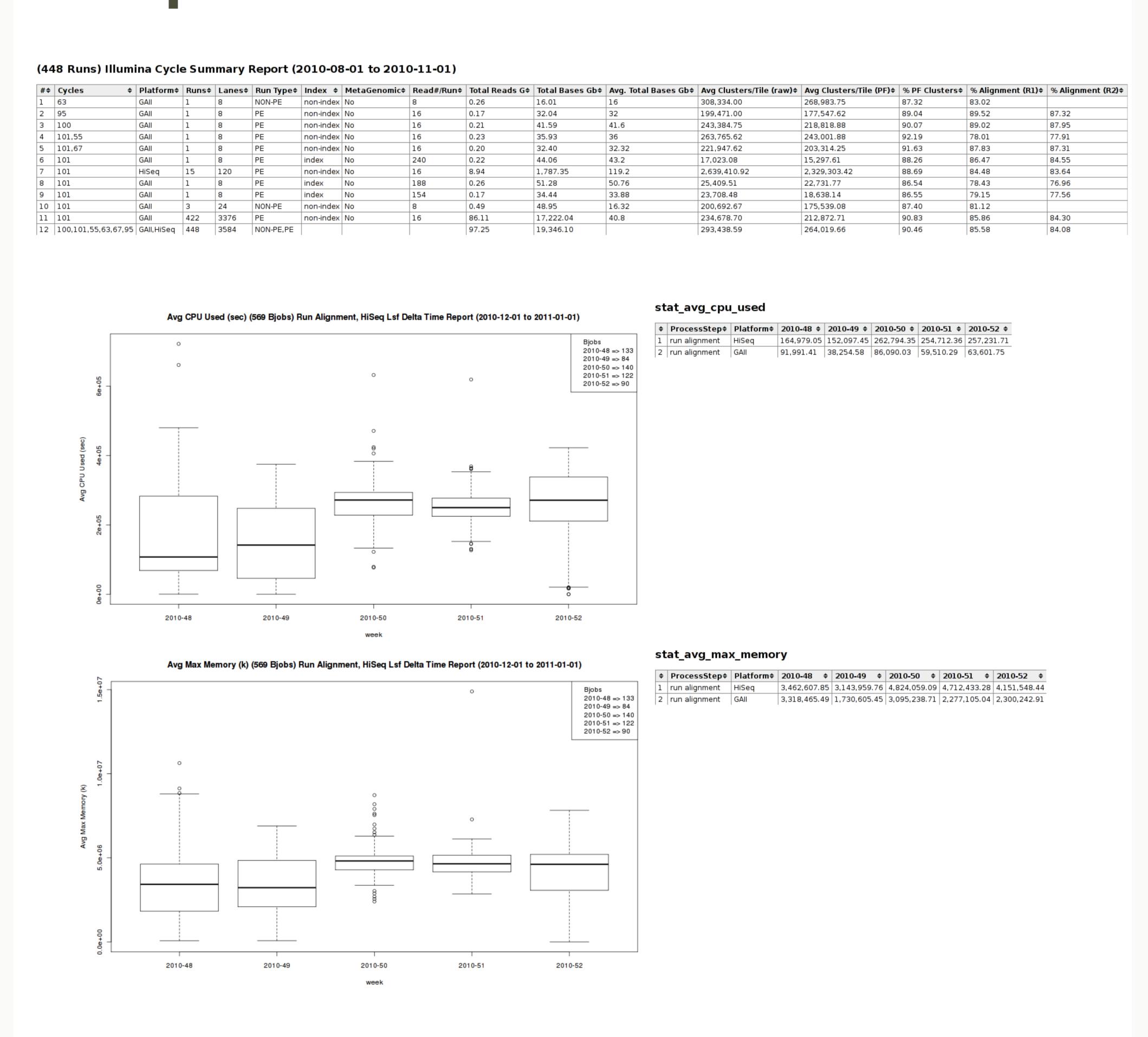


add recipies and assign to instruments

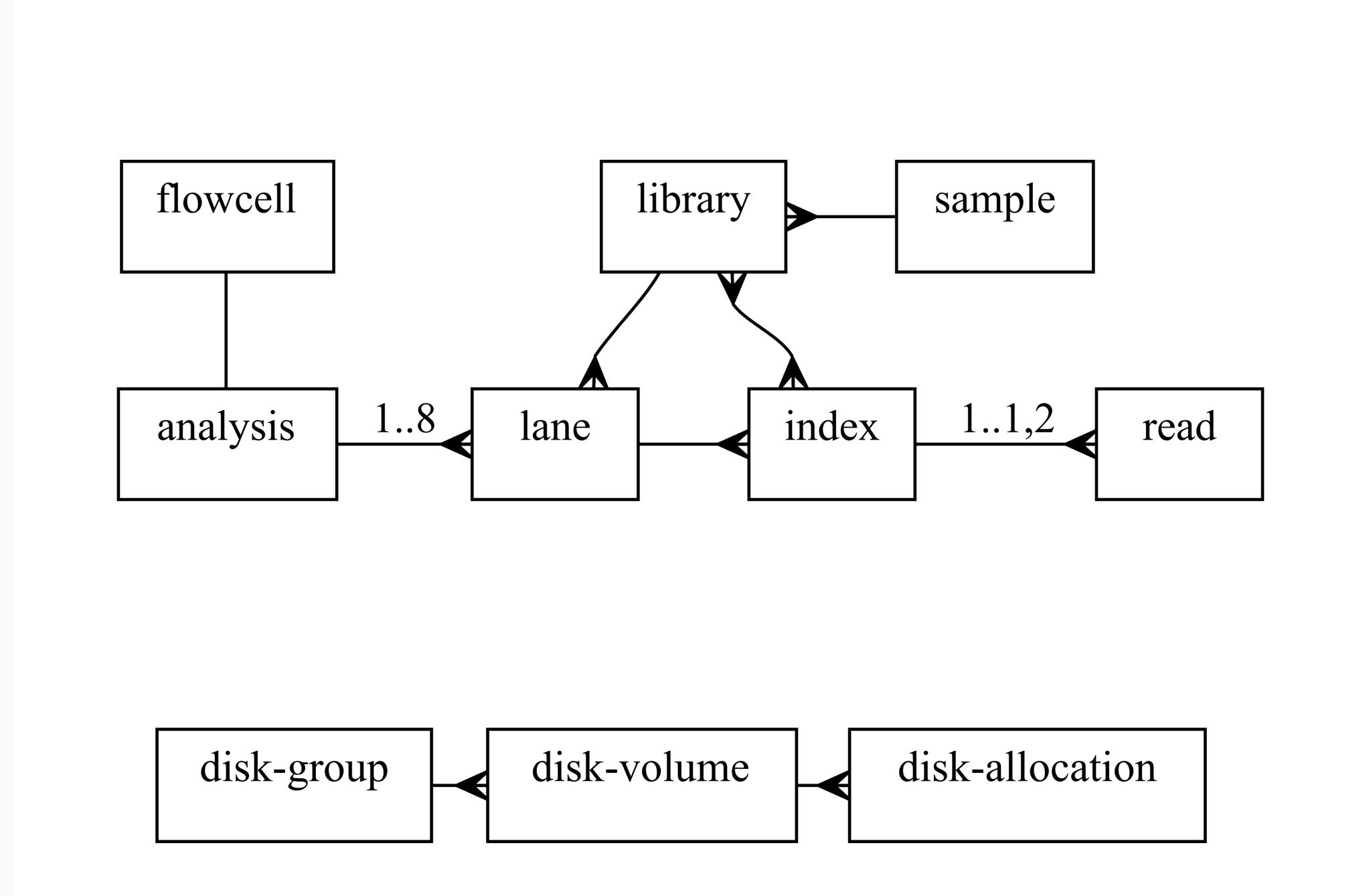
Disk Management



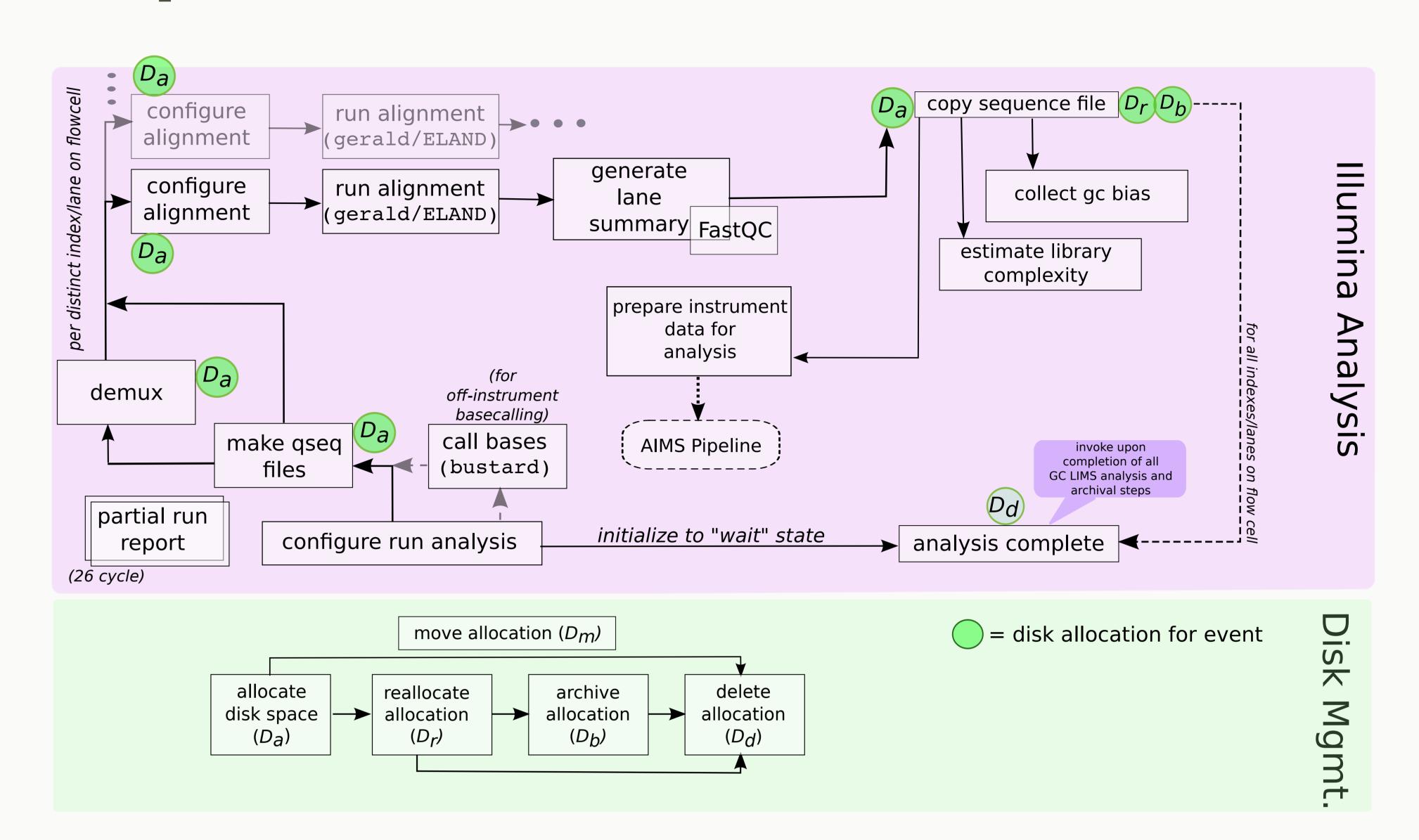
Reports

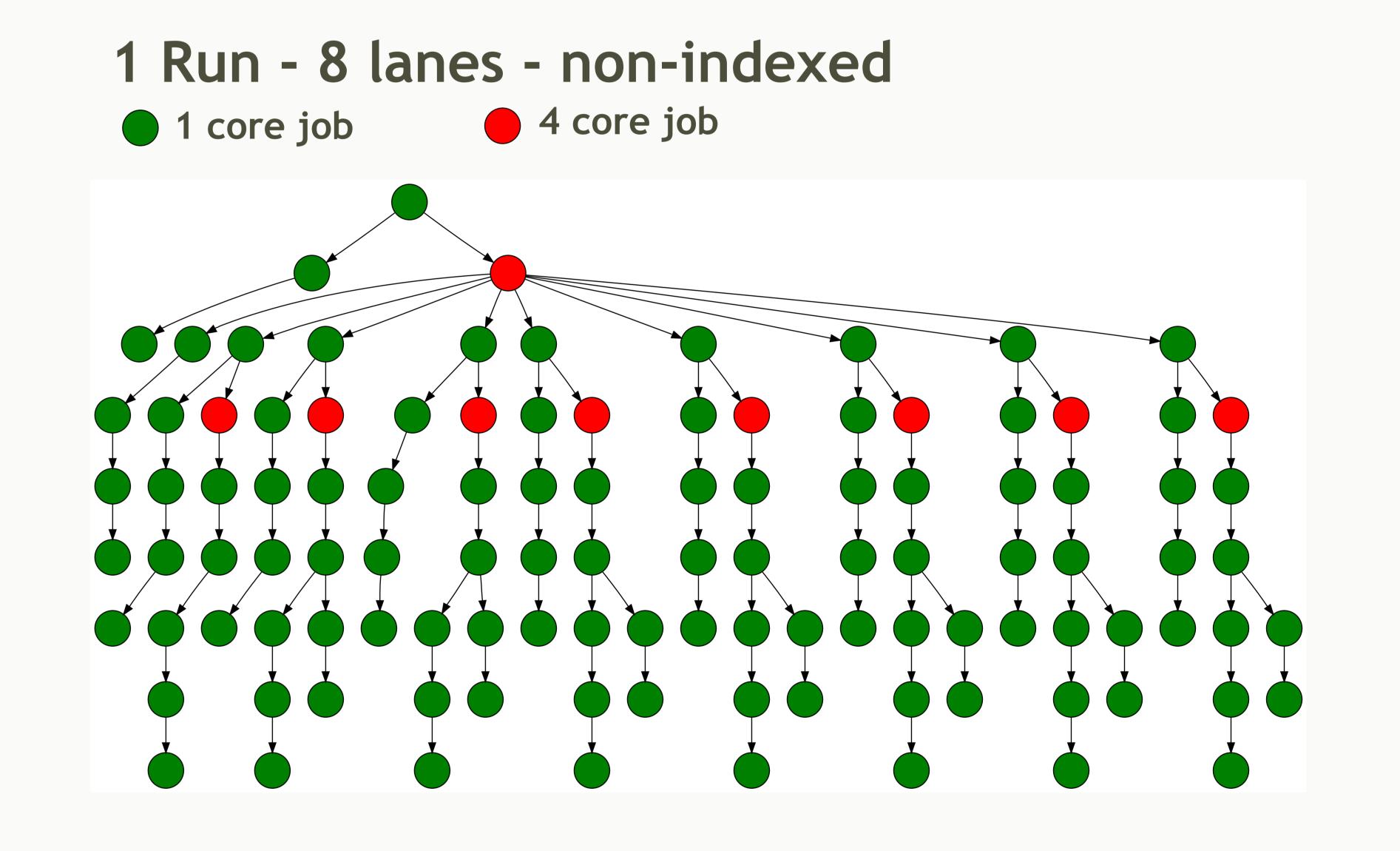


Schema



Pipeline





Future Work

- o Integrate CASAVA 1.8
- o Further analyze compute resources and optimize
- o Scale with increased number of HiSeq instruments



