SolexaTools: An Open Source Sequence Analysis Framework

Brian D. O'Connor, Jordan Mendler, and Stanley F. Nelson

The SolexaTools project was started to meet the computation infrastructure needs of scientists using the Solexa (now Illumina) massively parallel sequencing platform. This open source project provides two key features for the community. First, a laboratory information management system (LIMS) for tracking sequencing runs (SolexaLIMS) and, second, a pipeline framework for organizing the analysis of data (SolexaPipeline). These tools allow scientists to automate most of the processing, alignment, and report generation necessary when working with such large amounts of sequence data. This framework has been successfully deployed in our group and is a key part of all sequencing projects currently underway in our lab. In addition to technical details of the SolexaTools project, we also discuss the general information technology infrastructure requirements of next generation sequencing projects. SolexaTools is freely available under the GNU General Public License (GPLv3) and has been written to allow for easy customization and community driven development through the project website (http://solexatools.sourceforge.net).