

## **An Integrated Database And Viewer For Tracking Illumina and SOLiD Sequencing and Analysis**

Richard Varhol, Tina Wong, Irene Li, Margret Asgeirsdottir, Yaron Butterfield, Richard Corbett, Anthony Fejes, Greg Stazyk, Inanc Birol & Steven Jones

With over 4611 and 86 libraries sequenced with 17 Illumina and 11 SOLiD instruments respectively, an integrated auditing system has been developed to incorporate metadata from the Laboratory Information Systems (LIMS) with the monitoring of the different analysis pipelines developed within the GSC.

The quality of the raw reads is first assessed for a host of QC level-0 metrics some of which include contamination, sample swaps, nucleotide distribution, read frequency distribution and most frequent read types, prior to being stored in the database. Subsequent processing and analysis based on the outcome of the level-0 analysis is dependent on which one of the 24 different experimental protocols has been specified for a library and would include the merging of existing lanes or slides. Specific library analysis is triggered once the data has been merged, after which an assessment of the library, Level-1, is made and recorded.

To assist in the monitoring of the above-mentioned processes in addition to the recording of data dissemination to collaborators and sequencing repositories, we have developed an in-house tracking platform using Python, PostgreSQL and sqlalchemy to create specific language-independent application programming interfaces (APIs) through the use of XML-RPC, enabling developers of customized analysis pipelines to integrate with the tracking database.

A separate user-driven website interfaces with the tracking database displaying summary information from various aspects of the project.