Semantic Infrastructure Domain User Story 3

Semantic Infrastructure Services

- Master List
- Periodic Table of Semantic Infrastructure Services
- User Story Matrix

Semantic Infrastructure Domain User Story 3:
Automatically discover analytical steps for Illumina bead array analysis using inference based on the semantic metadata of the parameters.

Domain Description

The Illumina BeadChip is proprietary method of performing multiplex gene expression and genotyping analysis. The essential element of BeadChip technology is the attachment of oligonucleotides to silica beads. An informaticist is working with a cancer researcher to study expression profiles related to proto-oncogenes in T-cell leukemias. It is the first time either has worked with this technology, and the inforaticist is in the process of developing an analytical pipeline. He performs a search for such an analytical pipeline, and a number of steps that can be linked together are presented to him based on the Illumina bead output data type and his end goal of identifying gene annotations. The pipeline that was inferred using semantic metadata is: BeadArray-specific variance stabilization and gene annotation at the probe level. The informaticist knows he needs a quality control step at the beginning to determine whether an experiment run was successful or produced bad data. He searches "quality control" and "expression data" for analytical services, and finds an option specific to Illumina bead arrays. He also knows that the control and experimental runs will need to be normalized. The results from his search apply to gene expression matrices, so he will need a translation step. He enters the bead array format as the input and the gene expression matrix as the output and finds what he needs. Fortunately, the analytical step fits right in before probe annotation, which can also work on a gene expression matrix provided the bead identifiers are included. He saves the workflow, types in some notes about it, and shares it with his cancer researcher colleague.

Technical Description

The discovery of analytical steps utilizes inference over semantic annotations of input and output parameters. The researcher selects the metadata types that will be input to the pipeline and those that will be output from the pipeline. The inference engine performs discovery steps, chaining inputs to outputs in an expanding set until all options are exhausted or the resulting type matches. Furthermore, when specific analytical steps are queries for, full-text and concept-based metadata searches are performed in conjunction with output/input matching to provide the bet possible results. Workflows are saved as a set of steps or as a set of constrains upon which workflows are dynamically generated to meet scientific goals.

Cross Reference

Support caB2B Services to integrate data on grid

Forum Request
Requirements Input
Use Cases

Brain Tumor in silico study - Pathology and Radiology data models

Forum Request
Requirements Input
Use Cases

ICR IRWG Requirements

Forum Request
Requirements Input
Use Cases

Related Services