Loading Genomic VCF Files into i2b2



Project Summary

This community project extends the current i2b2 query functionality by providing the ability to query for genotyped subjects by specific annotations related to genetic variants. Also, new query widgets have been built in the i2b2 web client that leverages existing i2b2 infrastructure for querying large strings of text stored in the observation_blob field of the observation_fact table.

Our package provides a *starting point* and working example of our local implementation:

- Source code for ETL process of VCF files
- Example i2b2 ontology & metadata XML
- Example data representation & SQL scripts
- Web Client widgets to submit queries

Downloads

Filename	Size	Description
i2b2_genomics_project_v1.zip	425 kB	Complete package containing example genomics implementation

Notes

User Roles

Users must have the DATA_DEID (de-identified) role to be able to execute a text search on the observation_blob field.

Full-text indexing

The observation_blob field in the observation fact table needs to be full text indexed or the genomics search will not work.