Demo Data

To utilize the existing i2b2 LARGESTRING query features to query patients with genomic variants, the genotype data is stored in the observation_fact table as an observation_blob. The observation_blob is a full text indexed field in the i2b2 database, which allows complex query constructs. The i2b2 software supports full text searches using the observation_blob field and the *contains* SQL function.

Scripts to load sample data can be found inside the package in “Scripts for Sample Data” folder. The following scripts are provided for SQL Server, Oracle, and PostgreSQL:

- concept_dimension.table.insert.sql
- metadata_genomics.table.create.sql
- metadata_genomics.table.insert.sql
- observation_fact.table.insert.sql
- table_access.table.insert.sql

Example observation_fact table with new fact rows loaded for a sample patient:

Sample Genotype data in the observation_fact table			
CONCEPT_CD	**INSTANCE_NUM**	**VALTYPE_CD**	**OBSERVATION_BLOB**
SO:0001483	1	B	rs377573539,T_to_C,MIR6723,homozygous_ref_ref,upstream
SO:0001483	2	B	rs6429759,C_to_T,AGMAT,homozygous_alt_alt,intron
SO:0001483	3	B	rs2298948,T_to_C,GCFC2,heterozygous_ref_alt,intron
SO:0001483	4	B	rs12640778,C_to_T,LINC01060,heterozygous_ref_alt,intron
SO:0001483	5	B	rs1060583,G_to_A,NECAB1,heterozygous_ref_alt,3'UTR
SO:0001483	6	B	rs533612,A_to_G,SIK2,homozygous_alt_alt,3'UTR
SO:0001483	7	B	rs4983407,C_to_T,MTA1,homozygous_ref_ref,intron

When one queries by dbSNP rs identifier, an example SQL query looks like:

Example rs identifier SQL query

```
select count(distinct patient_num)
  from observation_fact
  where contains(observation_blob, 'FAM148 AND (stop_loss OR missense)')
```

When one queries by gene name, an example SQL query looks like:

Example gene name SQL query

```
select count(distinct patient_num)
  from observation_fact
  where contains(observation_blob, 'rs183605470 AND heterozygous')
```