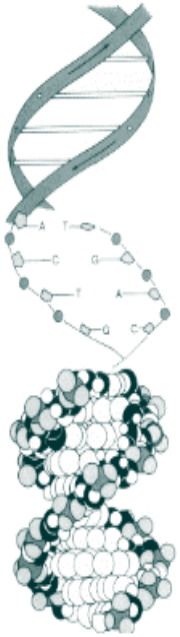


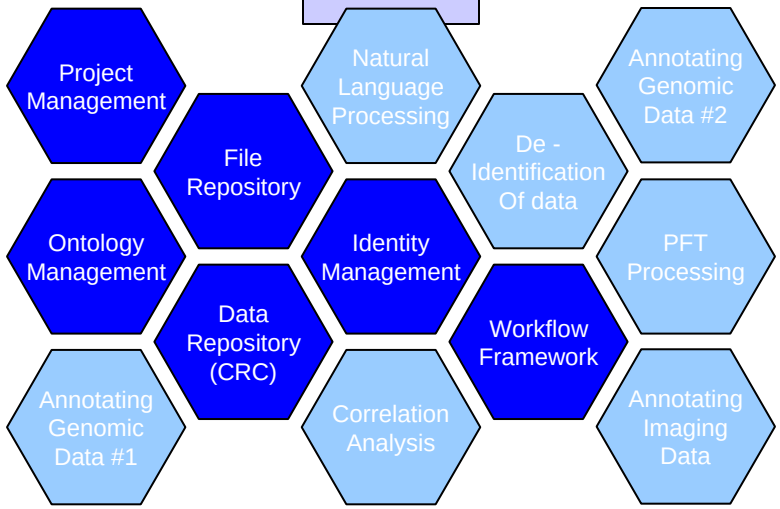
i2b2 Clinical Research Chart

Shawn Murphy MD, Ph.D.
Griffin Weber MD, Ph.D.
Michael Mendis
Vivian Gainer MS
Lori Phillips MS
Rajesh Kuttan
Wensong Pan MS
Henry Chueh MD
Susanne Churchill Ph.D.
John Glaser Ph.D.
Isaac Kohane MD, Ph.D.

The National Center for Biomedical Computing entitled Informatics for Integrating Biology and the Bedside (i2b2) Clinical Research Chart, what is it?

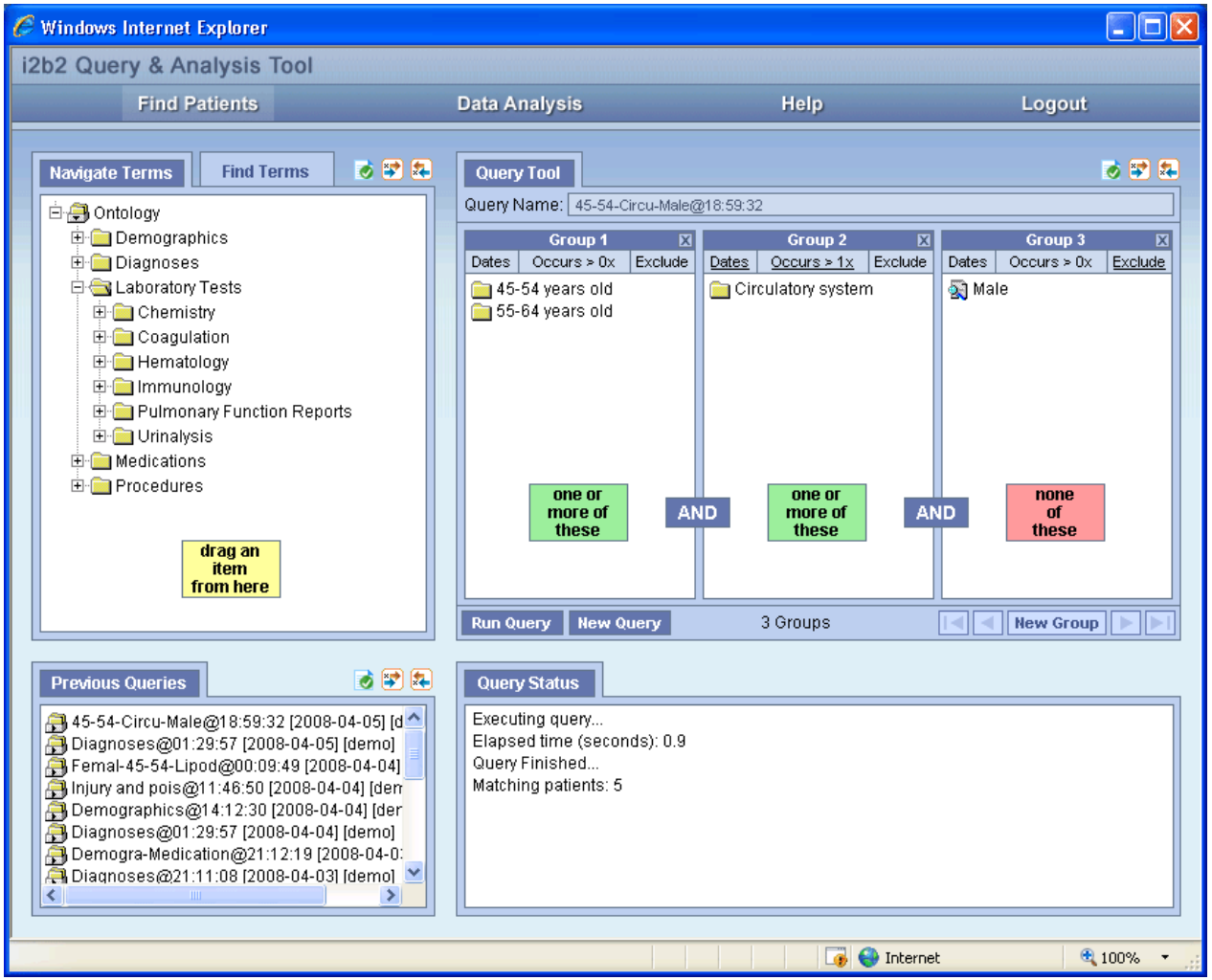
- Explicitly organized and transformed person-oriented clinical data optimized for clinical genomics research
- Formed as a collection of loosely coupled interoperable web services provided by i2b2 Cells
- An architecture that allows different studies to come together seamlessly
- An integration of clinical data, trials data, genotypic data, and knowledge annotation

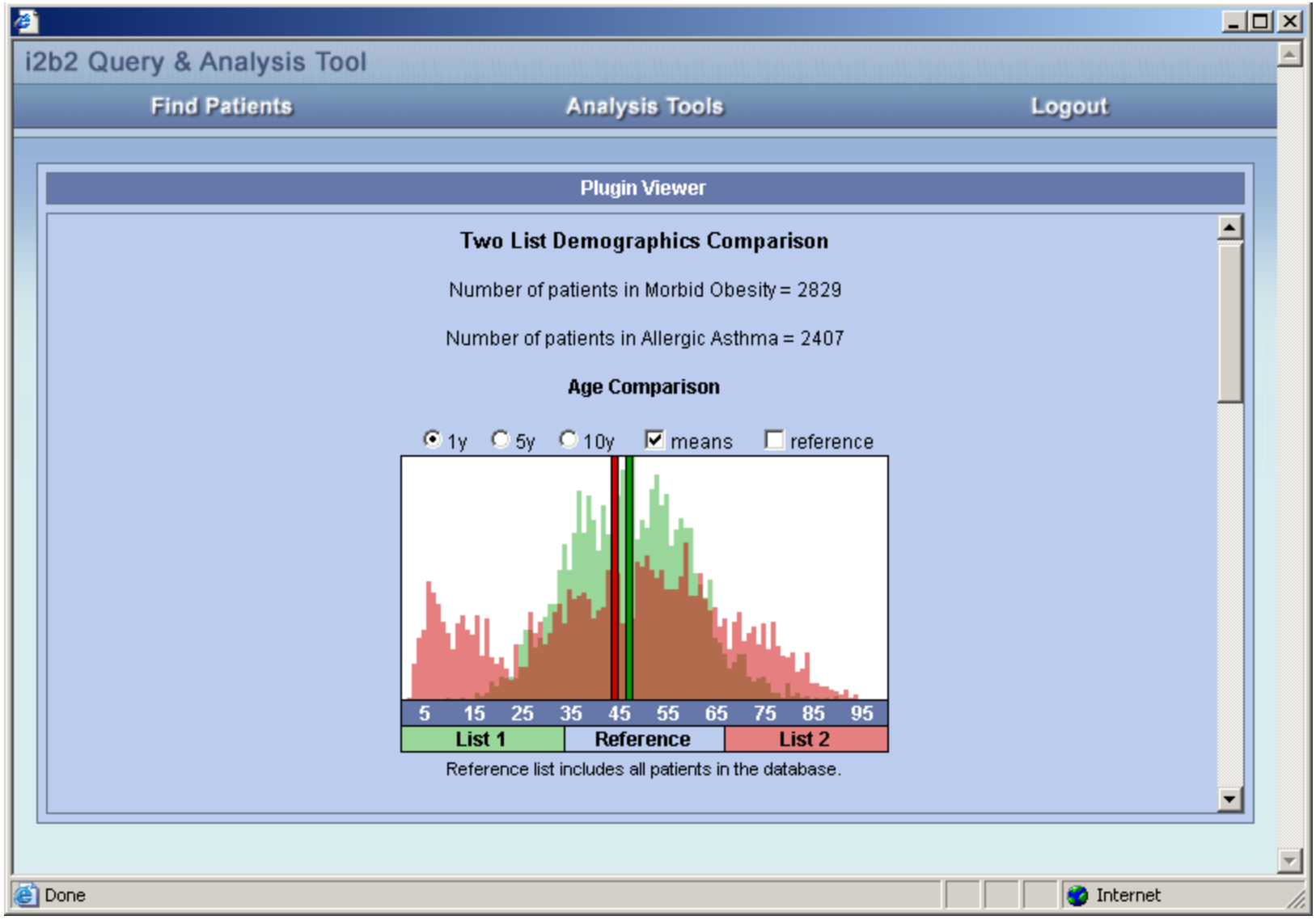
Enterprise-wide repurposing and distribution of medical record data for research

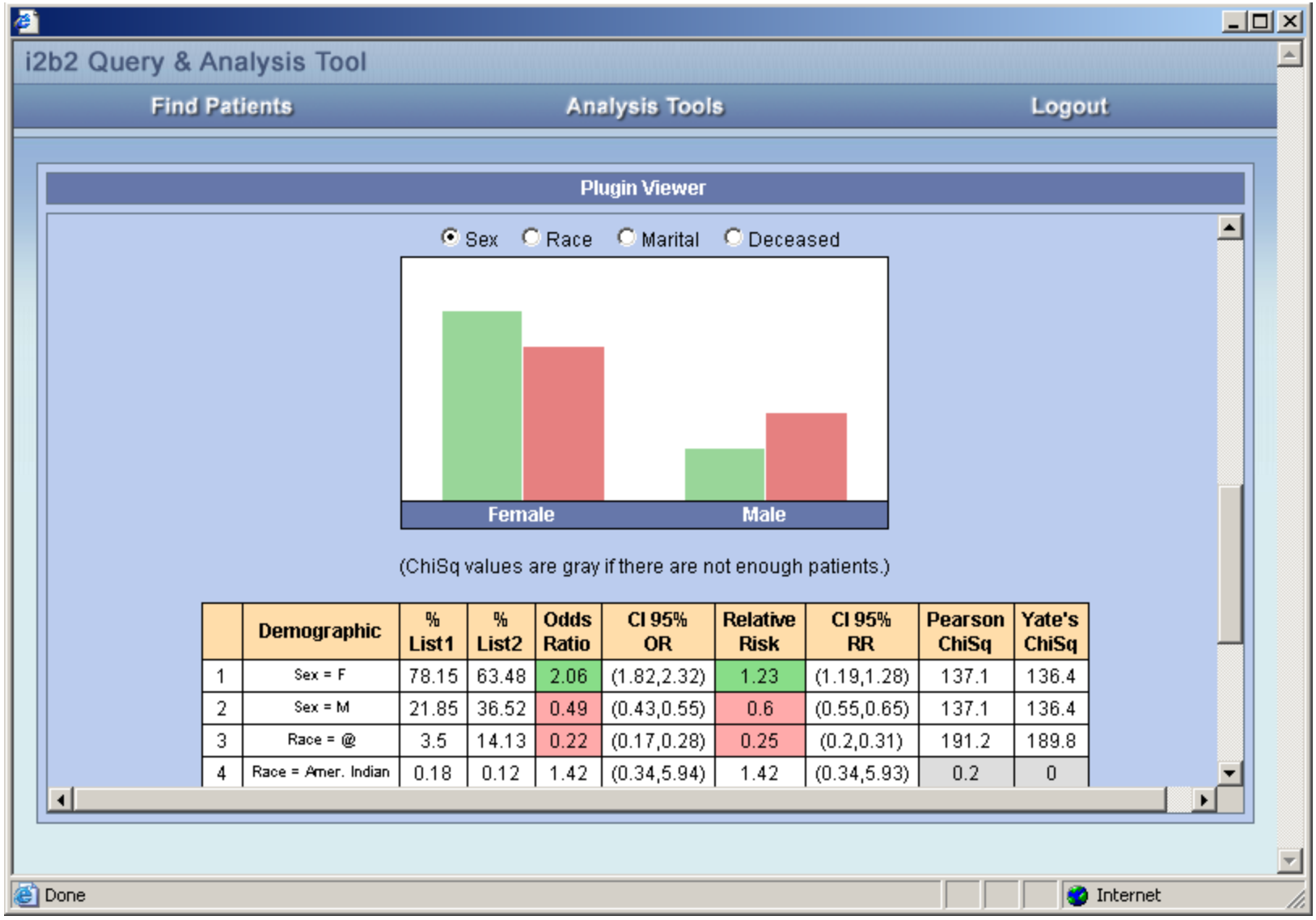


- Enable high performance collection of medical record data for querying and distribution
  - Enterprise web client
  - Create patient cohorts for further investigation
- Enable discovery within data on enterprise wide scale
  - Relevance networks

Enterprise web client



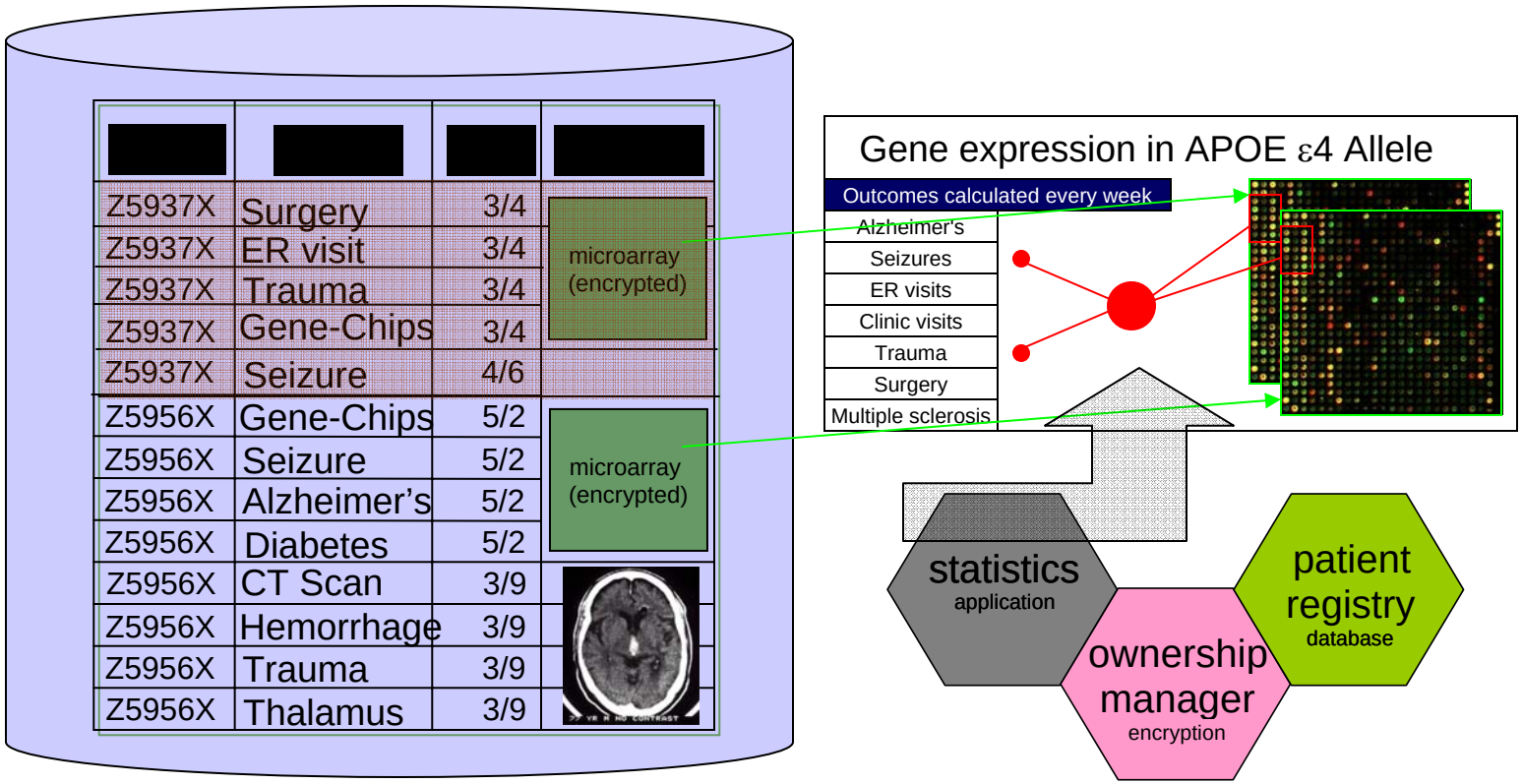

i2b2 Query & Analysis Tool
Find Patients
Analysis Tools
Logout
Plugin Viewer
Two List Demographics Comparison
Number of patients in Morbid Obesity = 2829
Number of patients in Allergic Asthma = 2407
Age Comparison
1y
5y
10y
means
reference
5
15
25
35
45
55
65
75
85
95
List 1
Reference
List 2
Reference list includes all patients in the database.
Done
Internet

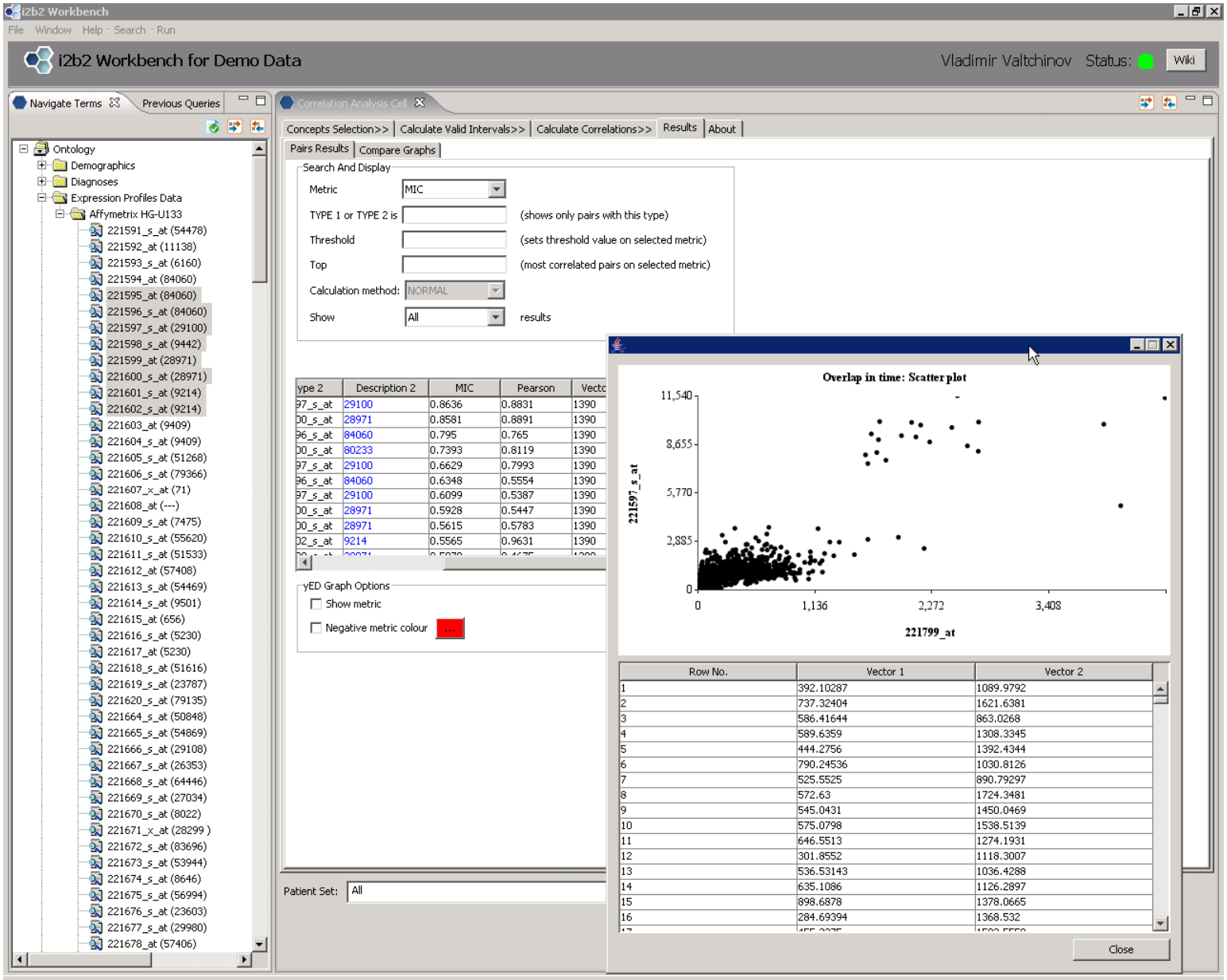


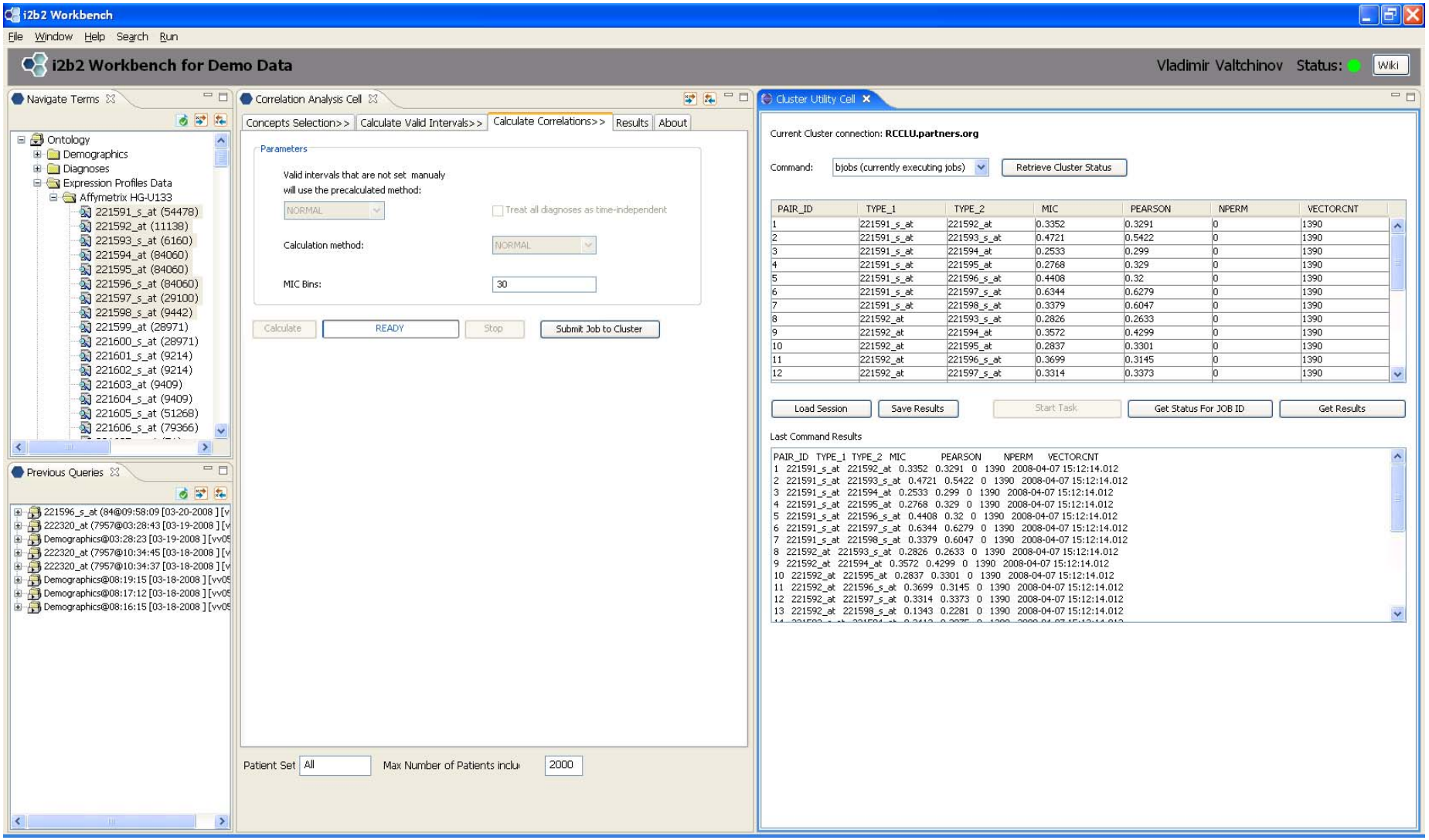
	Demographic	% List1	% List2	Odds Ratio	CI 95% OR	Relative Risk	CI 95% RR	Pearson ChiSq	Yate's ChiSq
1	Sex = F	78.15	63.48	2.06	(1.82,2.32)	1.23	(1.19,1.28)	137.1	136.4
2	Sex = M	21.85	36.52	0.49	(0.43,0.55)	0.6	(0.55,0.65)	137.1	136.4
3	Race = @	3.5	14.13	0.22	(0.17,0.28)	0.25	(0.2,0.31)	191.2	189.8
4	Race = Amer. Indian	0.18	0.12	1.42	(0.34,5.94)	1.42	(0.34,5.93)	0.2	0

Relevance Networks – unsupervised learning techniques

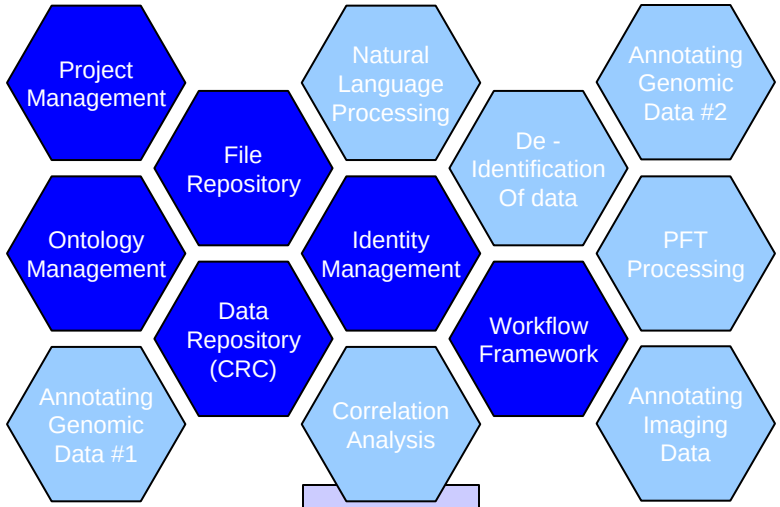
- Associates phenotypes and genotypes to generate novel hypothesis.




i2b2 Workbench
File Window Help · Search · Run
i2b2 Workbench for Demo Data
Vladimir Valtchinov Status:
Wiki
Navigate Terms
Previous Queries
Ontology
Demographics
Diagnoses
Expression Profiles Data
Affymetrix HG-U133
221591_s_at (54478)
221592_at (11138)
221593_s_at (6160)
221594_at (84060)
221595_at (84060)
221596_s_at (84060)
221597_s_at (29100)
221598_s_at (9442)
221599_at (28971)
221600_s_at (28971)
221601_s_at (9214)
221602_s_at (9214)
221603_at (9409)
221604_s_at (9409)
221605_s_at (51268)
221606_s_at (79366)
221607_x_at (71)
221608_at (---)
221609_s_at (7475)
221610_s_at (55620)
221611_s_at (51533)
221612_at (57408)
221613_s_at (54469)
221614_s_at (9501)
221615_at (656)
221616_s_at (5230)
221617_at (5230)
221618_s_at (51616)
221619_s_at (23787)
221620_s_at (79135)
221664_s_at (50848)
221665_s_at (54869)
221666_s_at (29108)
221667_s_at (26353)
221668_s_at (64446)
221669_s_at (27034)
221670_s_at (8022)
221671_x_at (28299)
221672_s_at (83696)
221673_s_at (53944)
221674_s_at (8646)
221675_s_at (56994)
221676_s_at (23603)
221677_s_at (29980)
221678_at (57406)
Correlation Analysis Cell
Concepts Selection>>
Calculate Valid Intervals>>
Calculate Correlations>>
Results
About
Pairs Results
Compare Graphs
Search And Display
Metric
MIC
TYPE 1 or TYPE 2 is
(shows only pairs with this type)
Threshold
(sets threshold value on selected metric)
Top
(most correlated pairs on selected metric)
Calculation method:
NORMAL
Show
All
results
ype 2
Description 2
MIC
Pearson
97_s_at 29100 0.8636 0.8831 1390
00_s_at 28971 0.8581 0.8891 1390
96_s_at 84060 0.795 0.765 1390
00_s_at 80233 0.7393 0.8119 1390
97_s_at 29100 0.6629 0.7993 1390
96_s_at 84060 0.6348 0.5554 1390
97_s_at 29100 0.6099 0.5387 1390
00_s_at 28971 0.5928 0.5447 1390
00_s_at 28971 0.5615 0.5783 1390
02_s_at 9214 0.5565 0.9631 1390
yED Graph Options
Show metric
Negative metric colour
Patient Set:
All
Overlap in time: Scatter plot
11,540
8,655
5,770
2,885
0
221597_s_at
0
1,136
2,272
3,408
221799_at
Row No.
Vector 1
Vector 2
1 392.10287 1089.9792
2 737.32404 1621.6381
3 586.41644 863.0268
4 589.6359 1308.3345
5 444.2756 1392.4344
6 790.24536 1030.8126
7 525.5525 890.79297
8 572.63 1724.3481
9 545.0431 1450.0469
10 575.0798 1538.5139
11 646.5513 1274.1931
12 301.8552 1118.3007
13 536.53143 1036.4288
14 635.1086 1126.2897
15 898.6878 1378.0665
16 284.69394 1368.532
Close

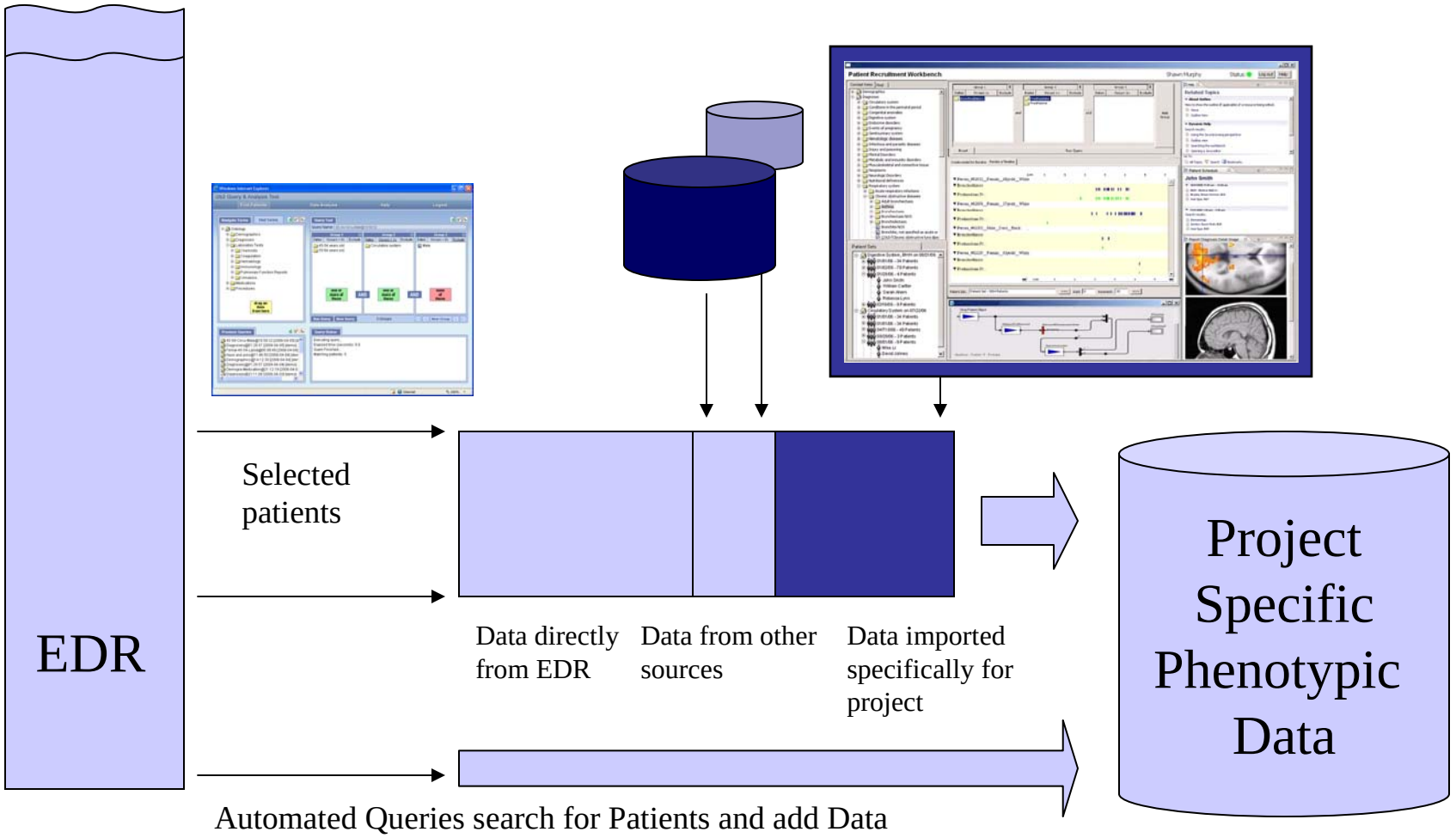

i2b2 Workbench
File Window Help Search Run
i2b2 Workbench for Demo Data
Vladimir Valtchinov Status:
Wiki
Navigate Terms
Ontology
Demographics
Diagnoses
Expression Profiles Data
Affymetrix HG-U133
221591_s_at (54478)
221592_at (11138)
221593_s_at (6160)
221594_at (84060)
221595_at (84060)
221596_s_at (84060)
221597_s_at (29100)
221598_s_at (9442)
221599_at (28971)
221600_s_at (28971)
221601_s_at (9214)
221602_s_at (9214)
221603_at (9409)
221604_s_at (9409)
221605_s_at (51268)
221606_s_at (79366)
Previous Queries
221596_s_at (84@09:58:09 [03-20-2008] [v
222320_at (7957@03:28:43 [03-19-2008] [v
Demographics@03:28:23 [03-19-2008] [vv0
222320_at (7957@10:34:45 [03-18-2008] [v
222320_at (7957@10:34:37 [03-18-2008] [v
Demographics@08:19:15 [03-18-2008] [vv0
Demographics@08:17:12 [03-18-2008] [vv0
Demographics@08:16:15 [03-18-2008] [vv0
Correlation Analysis Cell
Concepts Selection>>
Calculate Valid Intervals>>
Calculate Correlations>>
Results
About
Parameters
Valid intervals that are not set manualy
will use the precalculated method:
NORMAL
Treat all diagnoses as time-independent
Calculation method:
NORMAL
MIC Bins:
30
Calculate
READY
Stop
Submit Job to Cluster
Patient Set
All
Max Number of Patients inclu
2000
Cluster Utility Cell
Current Cluster connection: RCCLU.partners.org
Command:
bjobs (currently executing jobs)
Retrieve Cluster Status
PAIR_ID TYPE_1 TYPE_2 MIC PEARSON NPERM VECTORCNT
1 221591_s_at 221592_at 0.3352 0.3291 0 1390
2 221591_s_at 221593_s_at 0.4721 0.5422 0 1390
3 221591_s_at 221594_at 0.2533 0.299 0 1390
4 221591_s_at 221595_at 0.2768 0.329 0 1390
5 221591_s_at 221596_s_at 0.4408 0.32 0 1390
6 221591_s_at 221597_s_at 0.6344 0.6279 0 1390
7 221591_s_at 221598_s_at 0.3379 0.6047 0 1390
8 221592_at 221593_s_at 0.2826 0.2633 0 1390
9 221592_at 221594_at 0.3572 0.4299 0 1390
10 221592_at 221595_at 0.2837 0.3301 0 1390
11 221592_at 221596_s_at 0.3699 0.3145 0 1390
12 221592_at 221597_s_at 0.3314 0.3373 0 1390
Load Session
Save Results
Start Task
Get Status For JOB ID
Get Results
Last Command Results
PAIR_ID TYPE_1 TYPE_2 MIC PEARSON NPERM VECTORCNT
1 221591_s_at 221592_at 0.3352 0.3291 0 1390 2008-04-07 15:12:14.012
2 221591_s_at 221593_s_at 0.4721 0.5422 0 1390 2008-04-07 15:12:14.012
3 221591_s_at 221594_at 0.2533 0.299 0 1390 2008-04-07 15:12:14.012
4 221591_s_at 221595_at 0.2768 0.329 0 1390 2008-04-07 15:12:14.012
5 221591_s_at 221596_s_at 0.4408 0.32 0 1390 2008-04-07 15:12:14.012
6 221591_s_at 221597_s_at 0.6344 0.6279 0 1390 2008-04-07 15:12:14.012
7 221591_s_at 221598_s_at 0.3379 0.6047 0 1390 2008-04-07 15:12:14.012
8 221592_at 221593_s_at 0.2826 0.2633 0 1390 2008-04-07 15:12:14.012
9 221592_at 221594_at 0.3572 0.4299 0 1390 2008-04-07 15:12:14.012
10 221592_at 221595_at 0.2837 0.3301 0 1390 2008-04-07 15:12:14.012
11 221592_at 221596_s_at 0.3699 0.3145 0 1390 2008-04-07 15:12:14.012
12 221592_at 221597_s_at 0.3314 0.3373 0 1390 2008-04-07 15:12:14.012
13 221592_at 221598_s_at 0.1343 0.2281 0 1390 2008-04-07 15:12:14.012

- Repurpose medical record information for research studies
  - I2b2 Workbench
  - Natural language processing
- Enable genomic studies
  - Tissue/blood selection
  - Genetic data integration

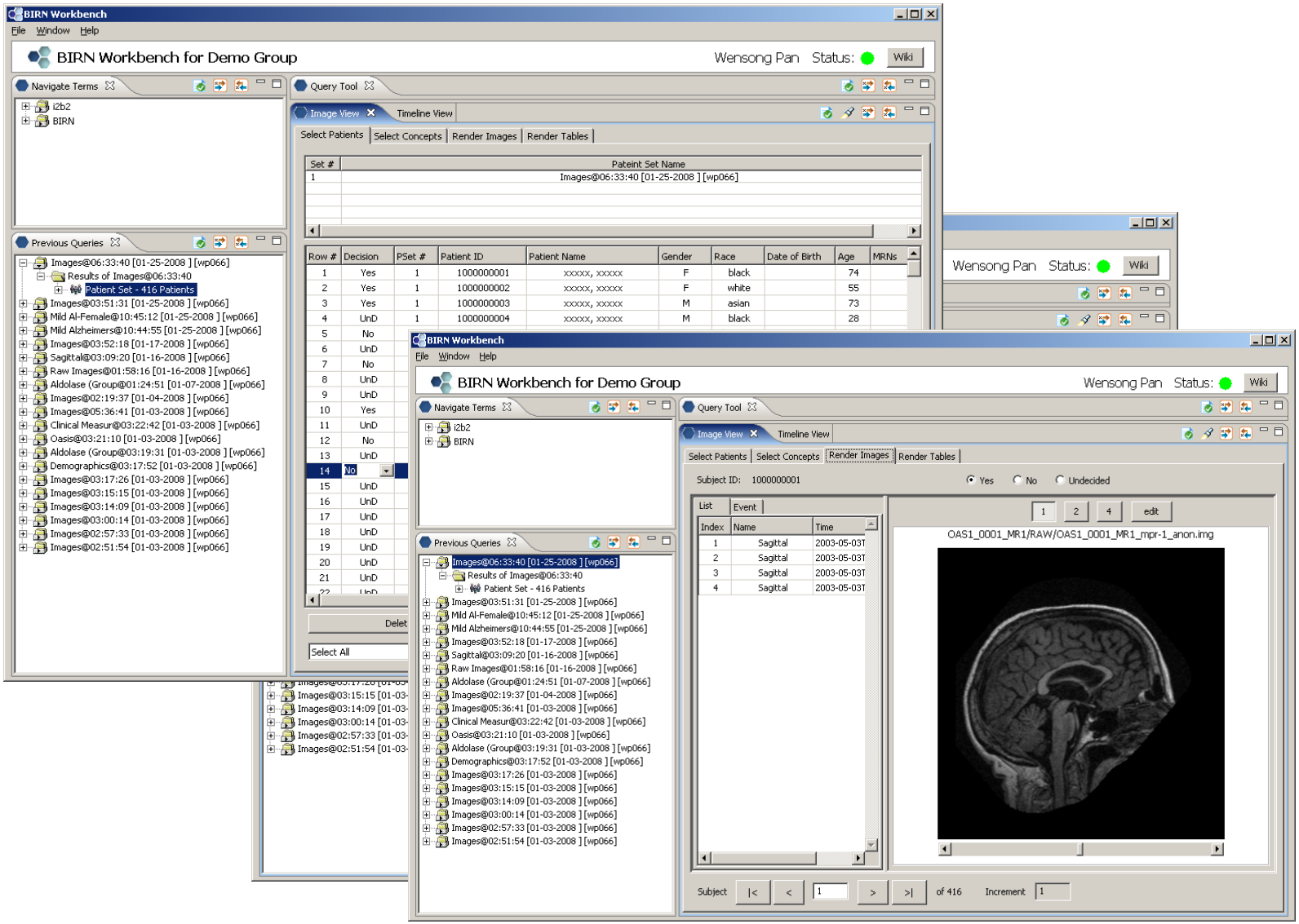


Use of medical record data in clinical studies focused upon genomics and pharmacology

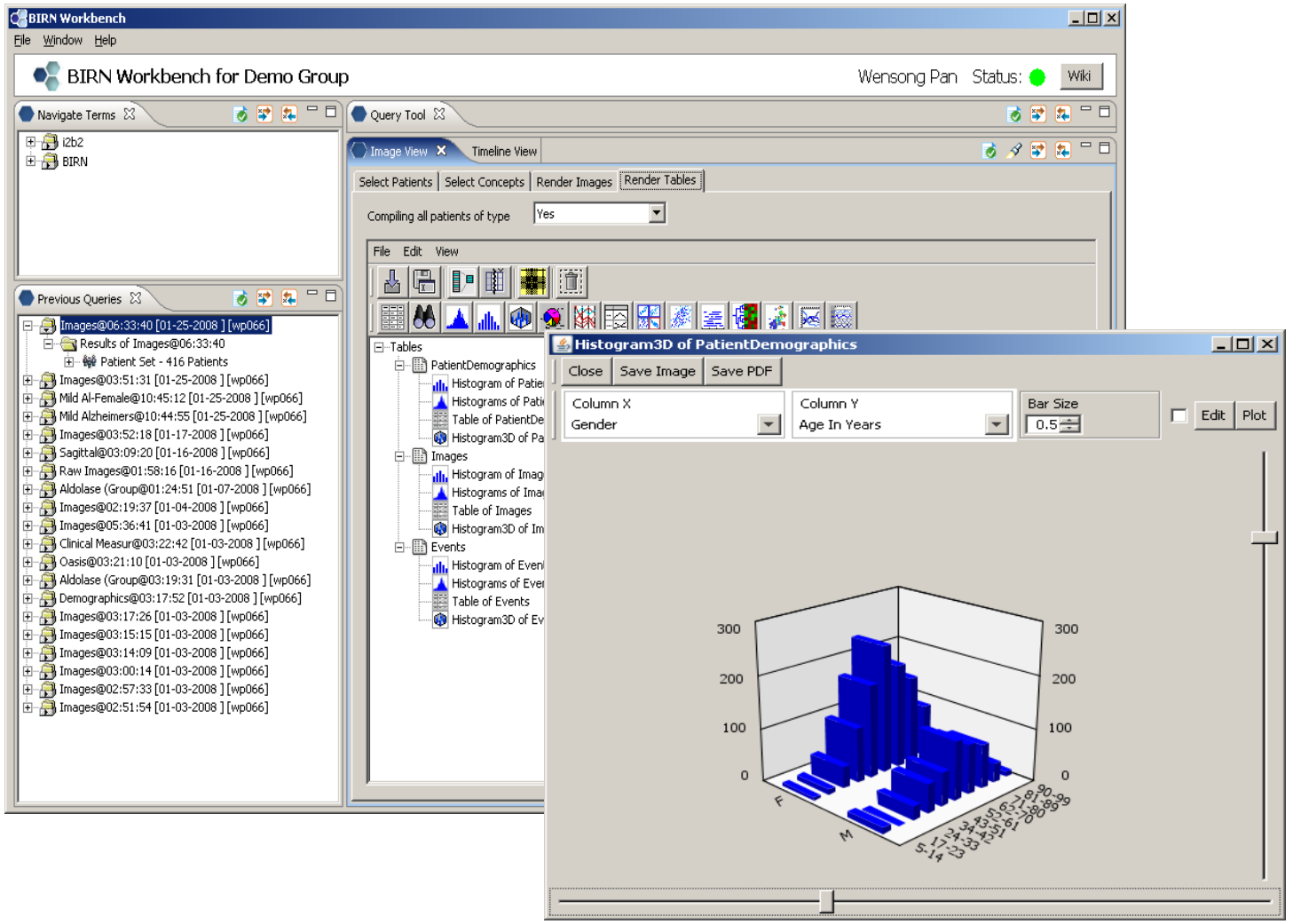
Set of patients is selected through Enterprise Repository and data is gathered into a data mart



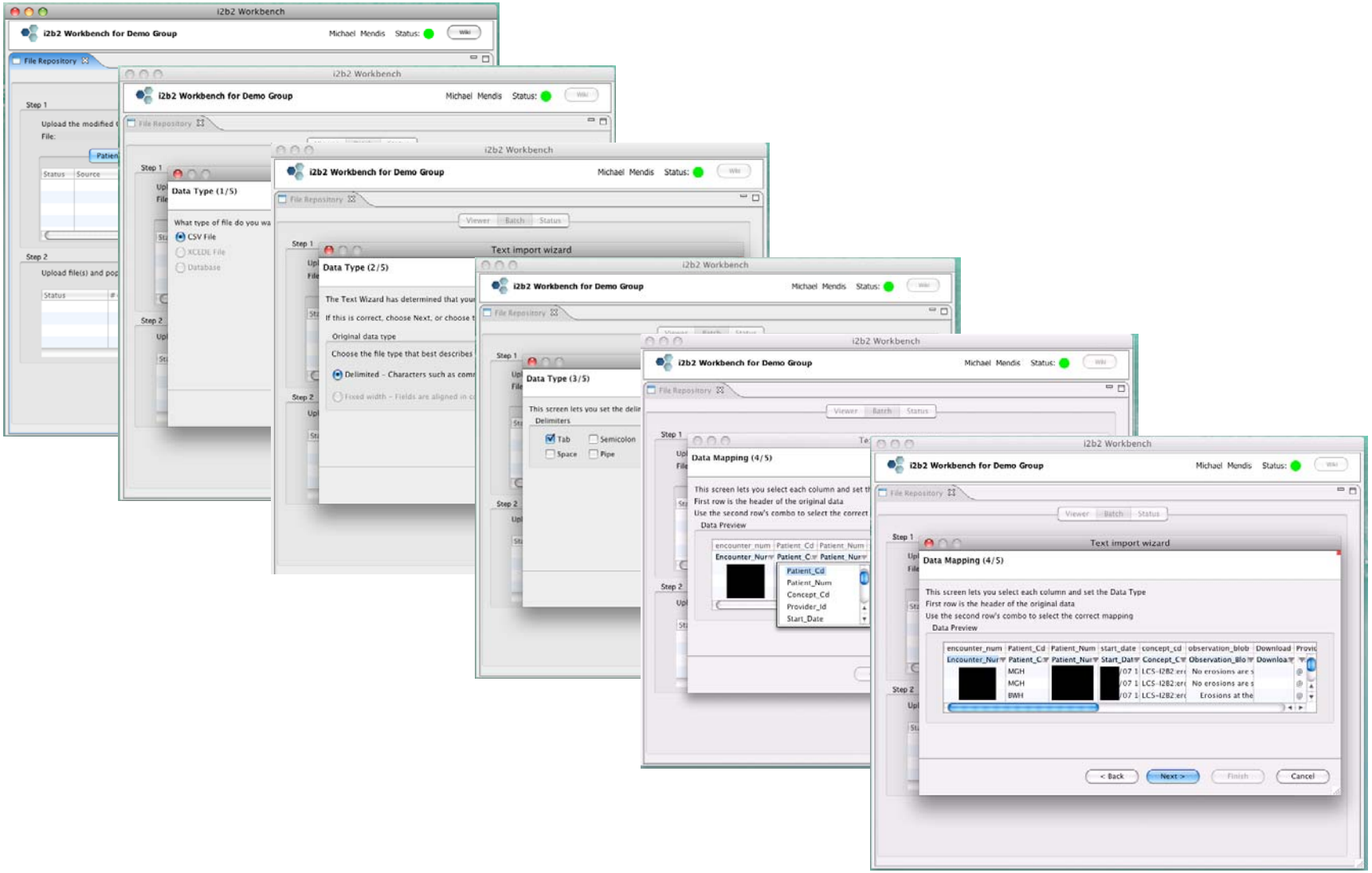
Patient population medical record review



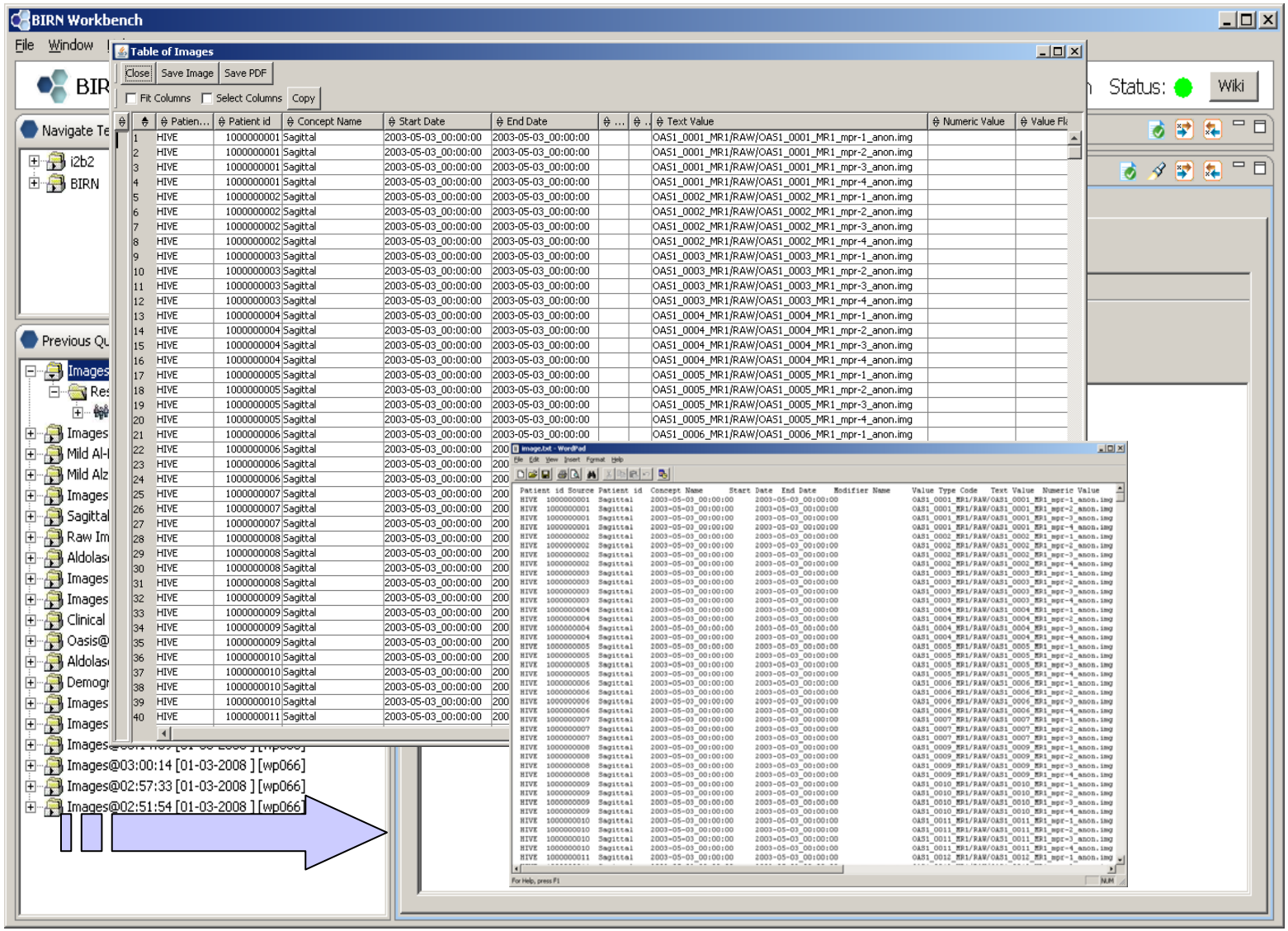
Exploration of targeted population



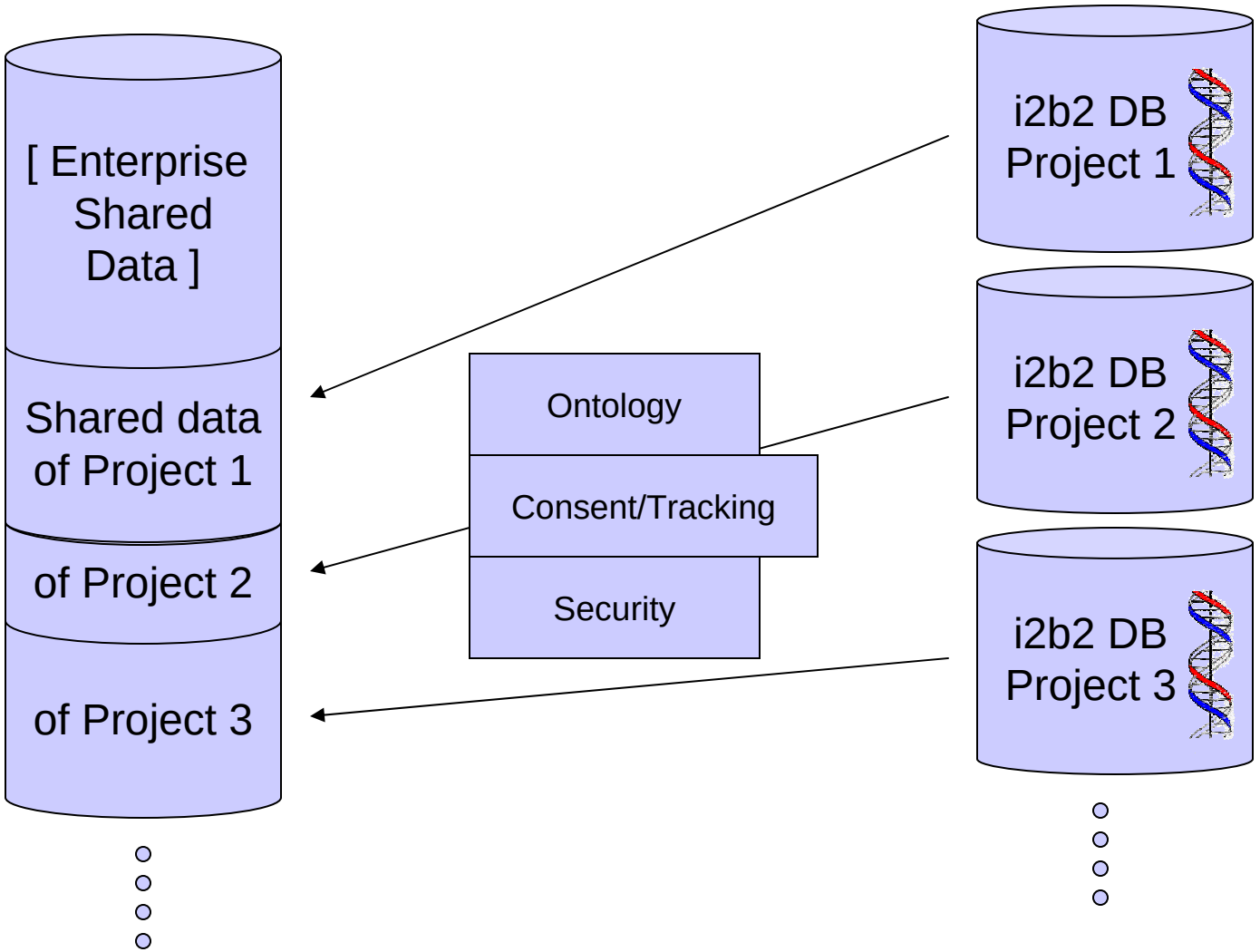
Import of batch data sets through i2b2 Workbench



Export of batch data sets for specific applications



Project data can be added back to Enterprise Repository



Enterprise Querytool allows:

- Selection of patient populations through complex queries
- Specifically targeted population analysis
- Privacy Management

Project Workbench allows:

- Patient population medical record review and exploration of data in a targeted population
- Use of Natural Language Processing
- Import of batch data sets
- Workflow support to simulate sequences of operations
- Export of batch outputs for specific applications
- Privacy management

Identity Management

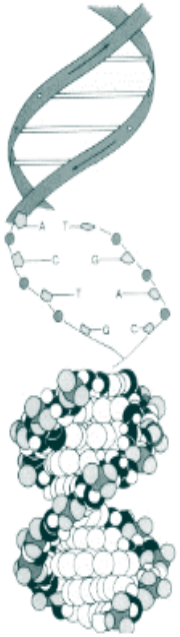
- Integration into existing medical record systems is an important requirement
- Levels of de-identification need to be supported
  - 1 – aggregate data only (obfuscate overly identifying items)
  - 2 – random records from a limited data set
  - 3 – full limited data set
    - 3.1 – de-identified text reports
  - 4 – fully identified data set
- Auditing to be tied into Project Management Cell
- Encryption key policy is an important difference between locals

Help outside of i2b2 and AUG

- Ontology management
  - NCBO
- Imaging Research
  - BIRN / XNAT
- Workflow management
  - [several options]
- Genomics integration
  - Vista
- Tissue management
  - Crimson

Organizing help within AUG

- New Web Site
- Issue Tracking
- Open Source Management, with a long term goal to help the following be supported:
  - - ETL
  - - Epidemiology workflows
  - - GWAS support



i2b2 Clinical Research Chart

Shawn Murphy MD, Ph.D.
Griffin Weber MD, Ph.D.
Michael Mendis
Vivian Gainer MS
Lori Phillips MS
Rajesh Kuttan
Wensong Pan MS
Henry Chueh MD
Susanne Churchill Ph.D.
John Glaser Ph.D.
Isaac Kohane MD, Ph.D.