

A Translational Engine at the National Scale: i2b2

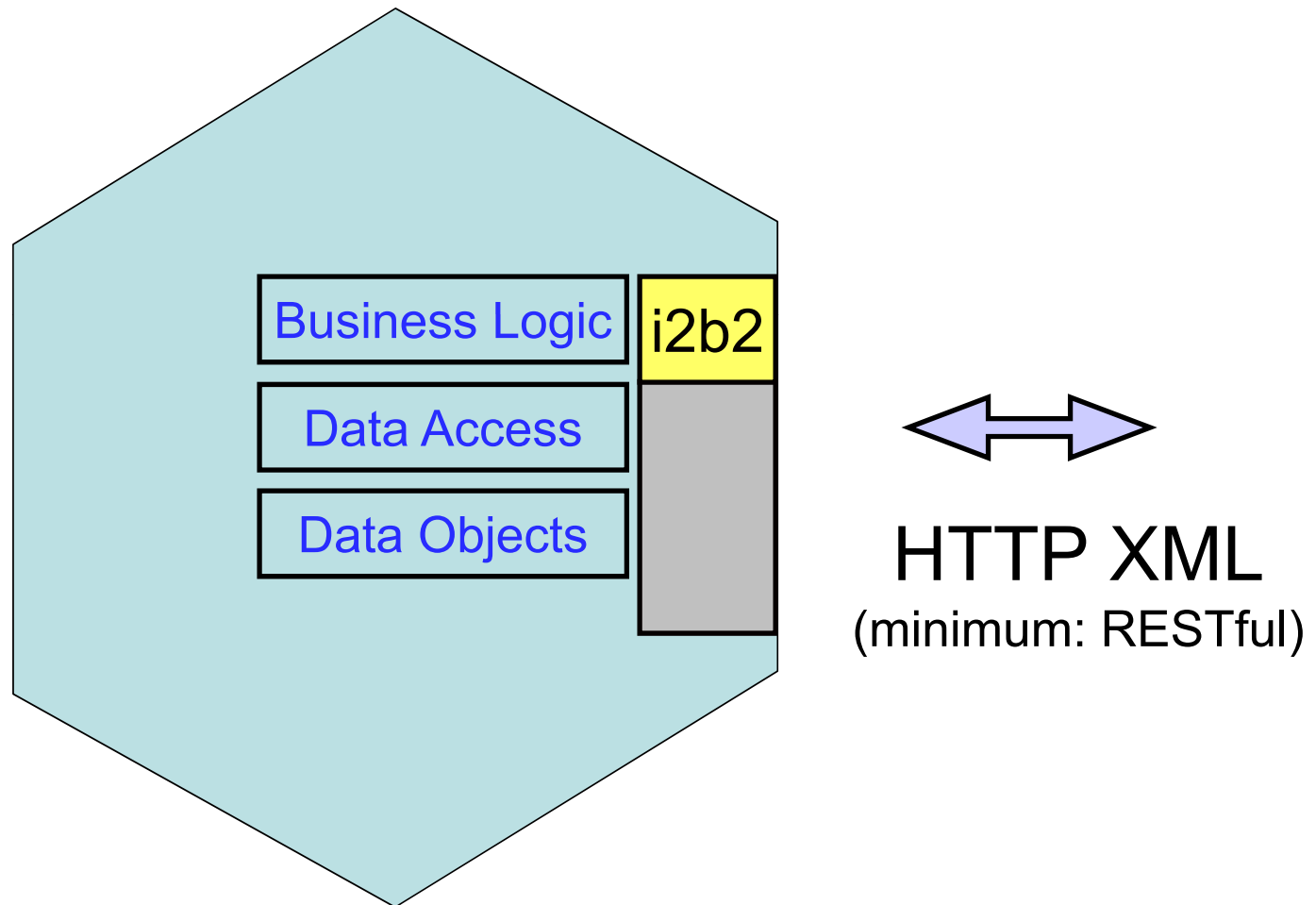
Shawn Murphy MD, Ph.D.



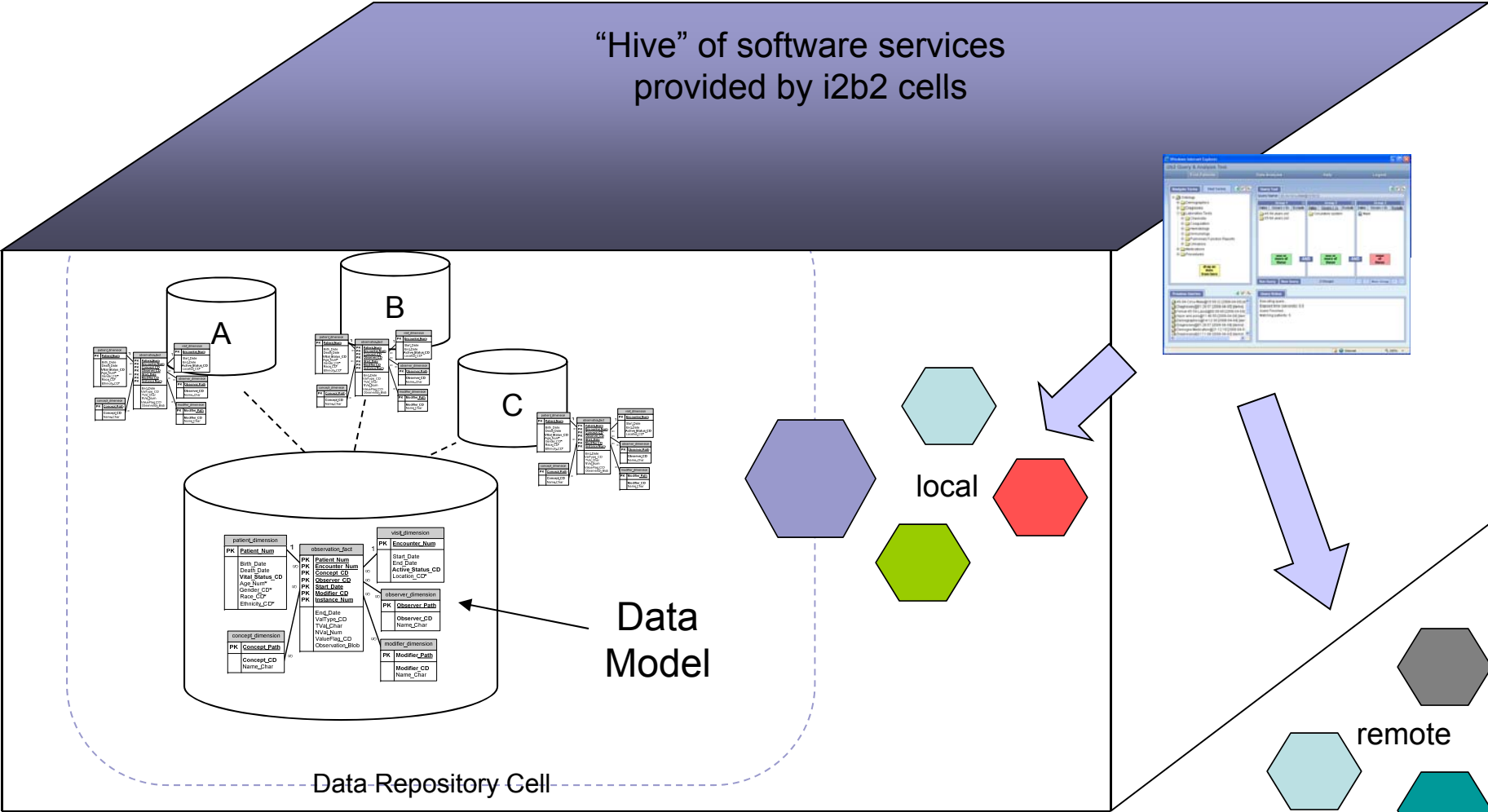
The National Center for Biomedical Computing entitled Informatics for Integrating Biology and the Bedside (i2b2), what is it?

- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
 - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
 - Software is built in a modular pattern that allows additions without disturbing core parts
 - Available as open source at <https://www.i2b2.org>

i2b2 Cell: The Canonical Software Module



An i2b2 Environment (the Hive) is built from i2b2 Cells



I2b2 Software components are distributed as open source

i2b2: Informatics for Integrating Biology & the Bedside - Windows Internet Explorer

Address bar: <https://www.i2b2.org/software/index.html>

Navigation: File Edit View Favorites Tools Help

Search: Google Search More >> Sign In Convert Select

Left sidebar menu:

- Archived Source Code
- Contributed
- Tutorial
- Guestbook *
- Statistics *

Main content area:

i2b2 Web Client Launch the AJAX web client in your web browser

Documentation

Hover over the modules below for the latest documentation:

Key: ■ i2b2 Core Cell ■ i2b2 Optional Cell ■ Workben ■ Web Client ■ CRC Plug

Diagram components:

- Project Management
- Natural Language Processing
- File Repository
- PFT Processing
- Ontology Management
- Identity Management
- Data Repository (CRC)
- Workflow Framework
- High Performance Computing Plug-in
- Correlation Analysis Plug-in
- Export Data Plug-in
- Table View Plug-in
- Text Analyzer Plug-in
- Import Data Plug-in
- i2b2 Workbench Application
- Annotator Plug-in

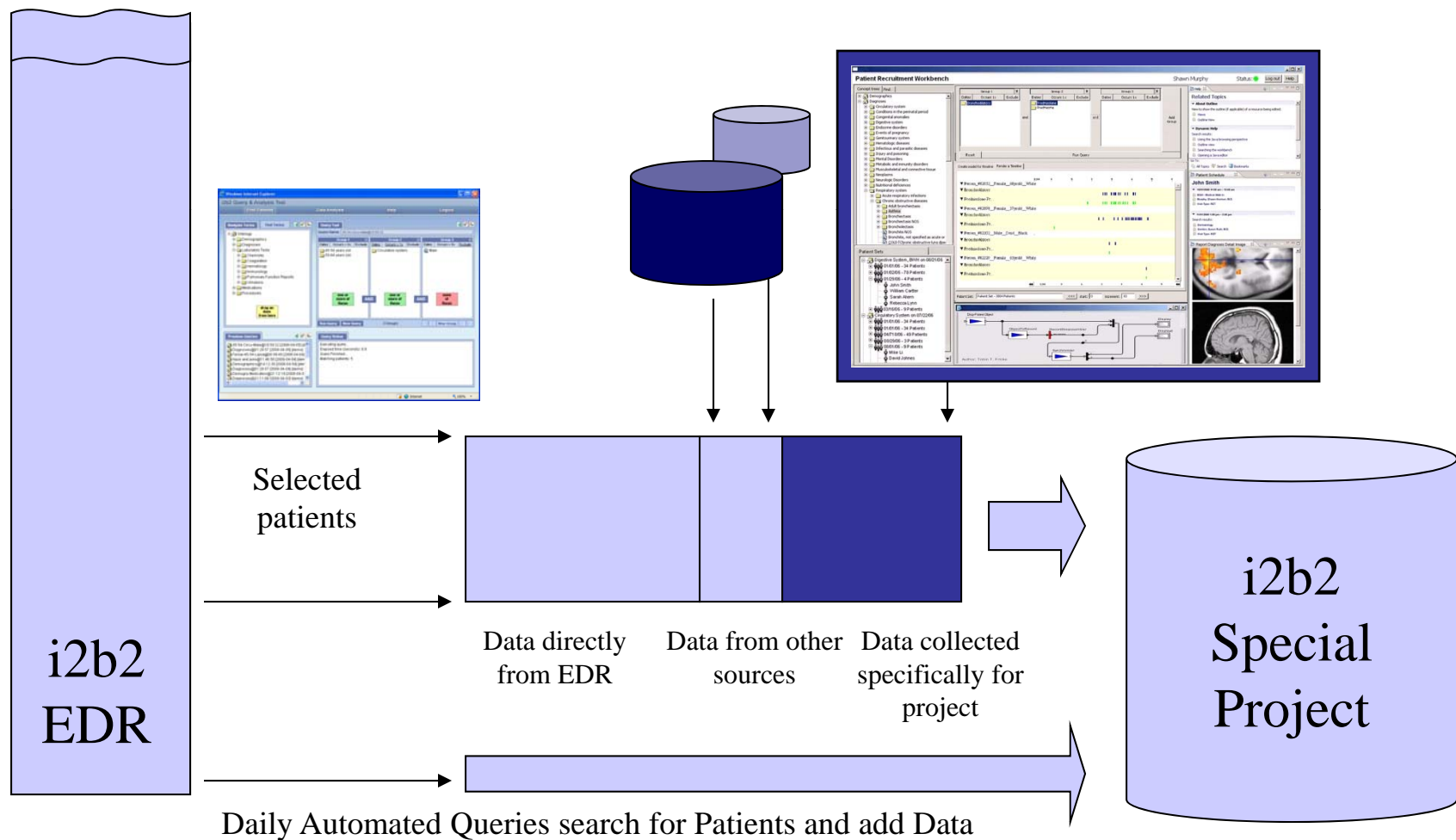
Tooltip for **i2b2 Workbench**:

The i2b2 Workbench is a collection of client-side components designed as Eclipse-based java plug-ins that communicate with i2b2 Cells and allow the investigator to query, analyze, and display the data of the hive, generally in greater depth than the web client.

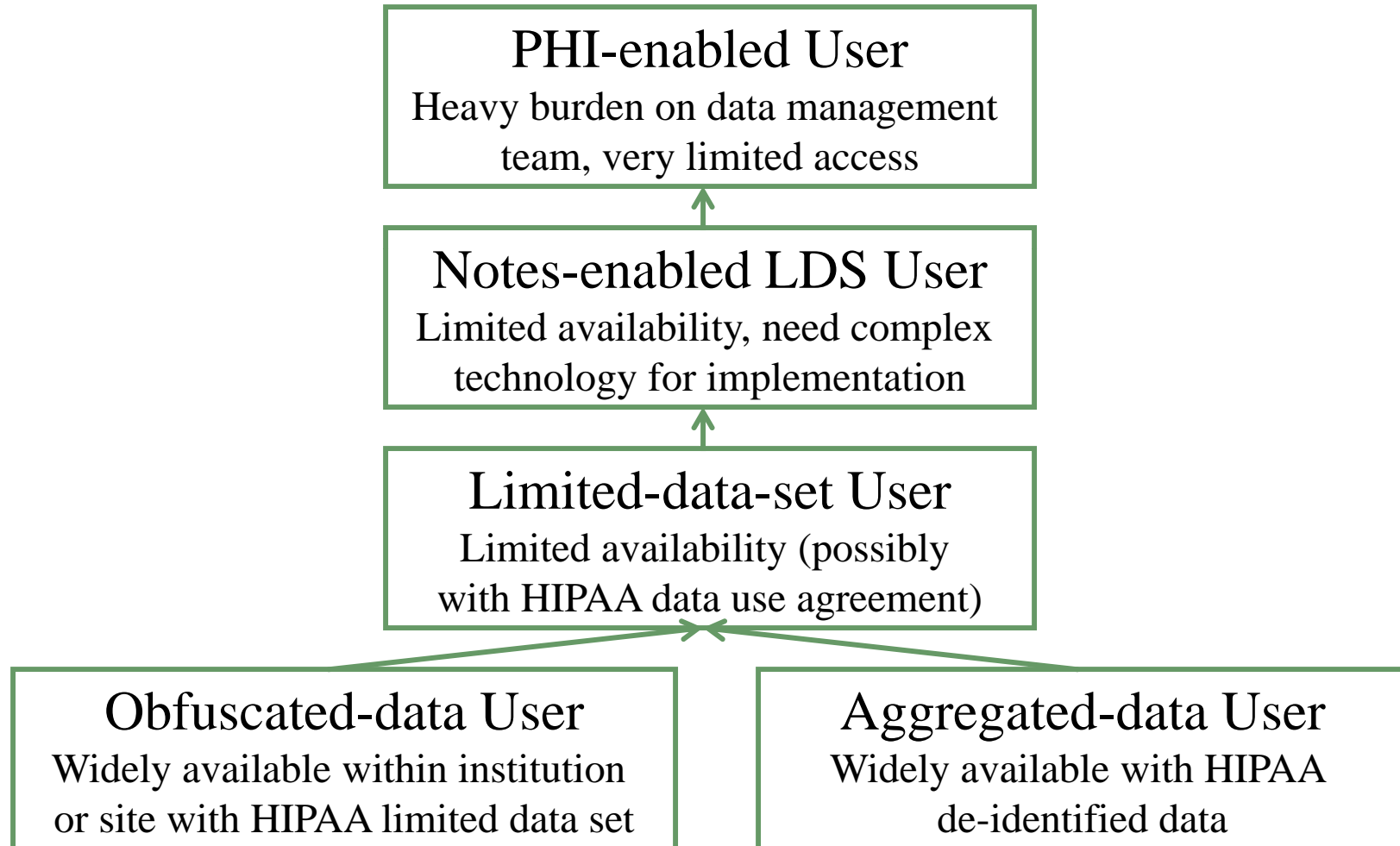
- Installation Guide
- Tutorial Document
- Developer's Guide
- Go to Download Client
- Go to Download Source

Footer: <https://www.i2b2.org/software/index.html#> Internet 100%

Perform deep studies with patient sets selected from Enterprise Data Repository



Privacy Levels in i2b2



Can We Trust the Phenotypes?

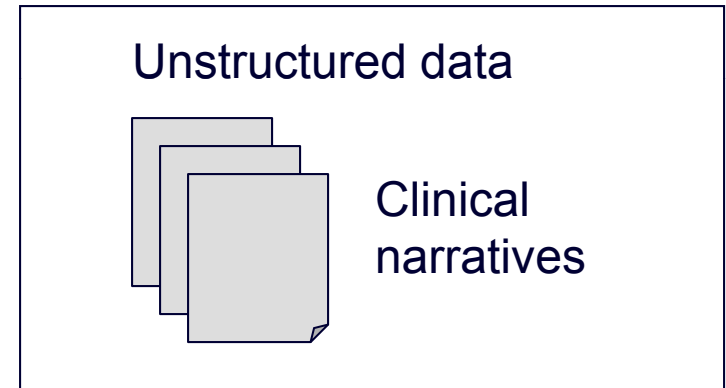
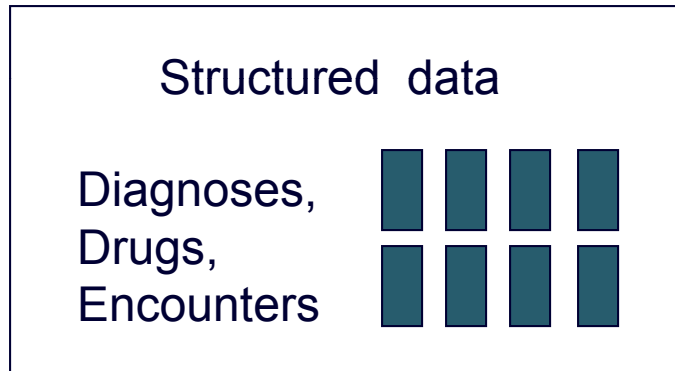
Validation Study (N = 185)

- Evaluate case and control algorithms compared to gold standard of diagnostic interview by expert clinician
- Recruit cases and controls as defined by informatics algorithm
- Interview by clinicians blinded to ascertainment group
- Recruited patients with depression or schizophrenia to enhance blinding

Jordan Smoller MD, ScD and team



Sensitivity vs. Specificity



Sensitivity



Specificity

Natural Language Processing

PROGRAMMER'S FILE EDITOR [D:\50210_1620\MiniDem1.txt]

File Edit Options Template Execute Macro Window Help

SOCIAL HISTORY: The patient is married with four grown daughters, **uses tobacco**, has wine with dinner. **Smoker**

PRINCIPAL DIAGNOSIS: LEFT LOWER LOBE PNEUMONIA

SECONDARY SOCIAL HISTORY: The patient is a **nonsmoker**. No alcohol. **Non-Smoker**

HISTORY SOCIAL HISTORY: **Negative for tobacco**, alcohol, and IV drug abuse.

PAST MEDICAL HISTORY: (1) Hip fracture. (2) Bronchiectasis.

BRIEF RESUME OF HOSPITAL COURSE:
63 yo woman with COPD, **50 pack-yr tobacco (quit 3 wks ago)**, **Past Smoker**

ALLERGIES: (1) Aspirin. (2) Ciprofloxacin. (3) Penicillin.

SOCIAL HISTORY: The patient lives in rehab, married. **Unclear smoking** history **???**
from the admission note...

PHYSICAL EXAMINATION: Temperature 37.2, pulse 88, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils

LABORATORY DATA: Sodium 148, potassium 3.4, chloride 97, bicarbonate

HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of **Lactobacillus** acidophilus to attempt to re-populate her gut. **Hard to pick**

HOSPITAL COURSE: The patient was seen and evaluated by the

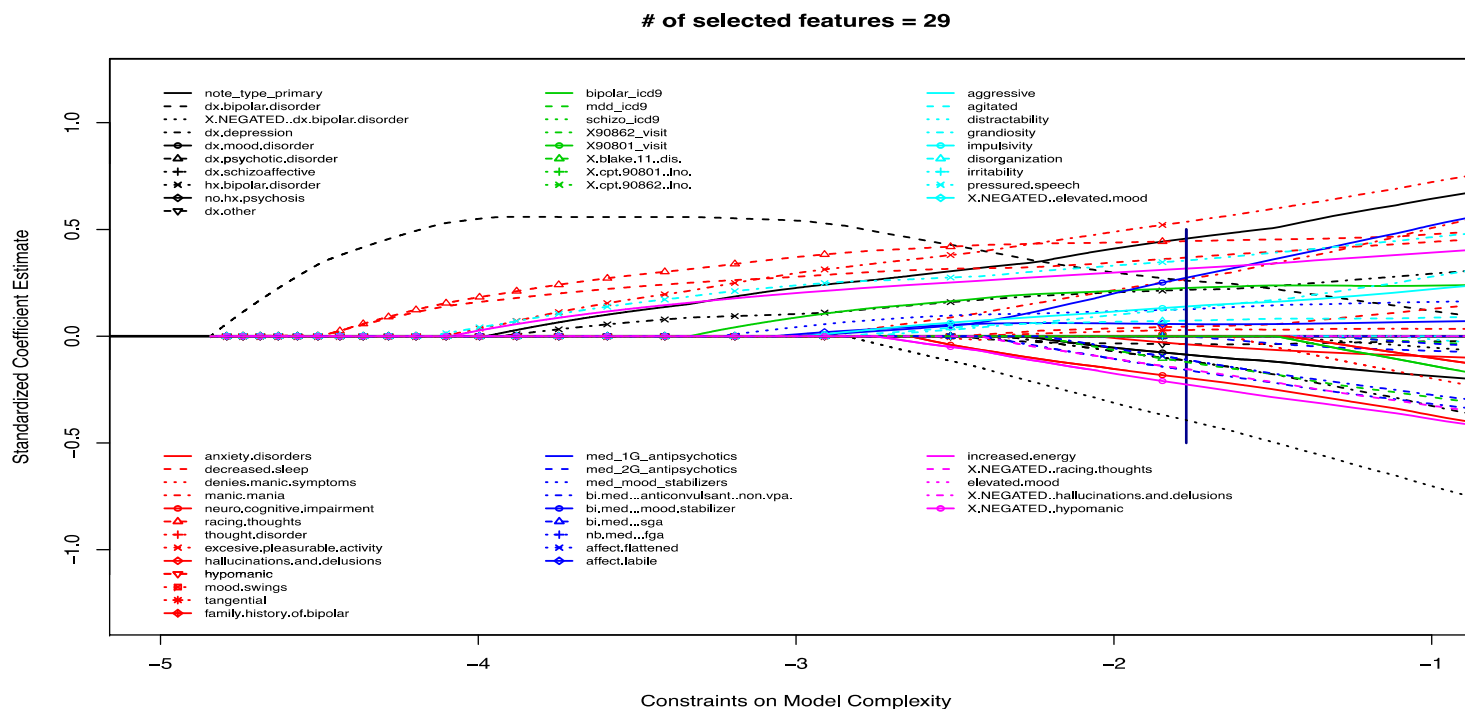
The patient SH: widow, lives alone, 2 children, no **tob/alcohol**. **Hard to pick**

Ln 44 Col 1 | 274 | WPT | [Esc Off] [No Wrap] [DOS] [INS] [NUM]

Train classification algorithms

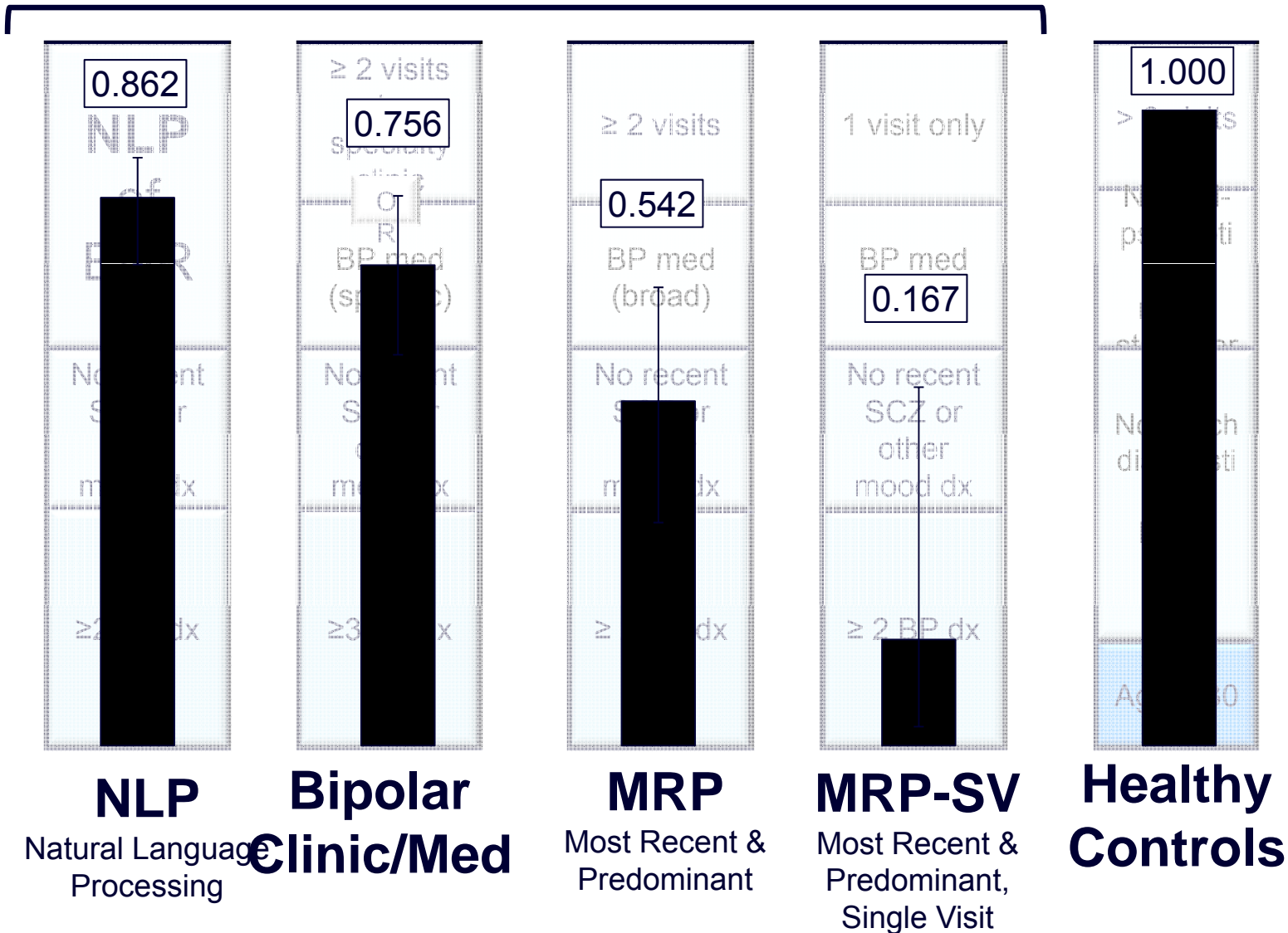
1. Over 300 words/phrases (features) were identified using chart review
2. Important features were selected for model using adaptive LASSO shrinkage

Tianxi Cai PhD and team



Bipolar Cohort Ascertainment

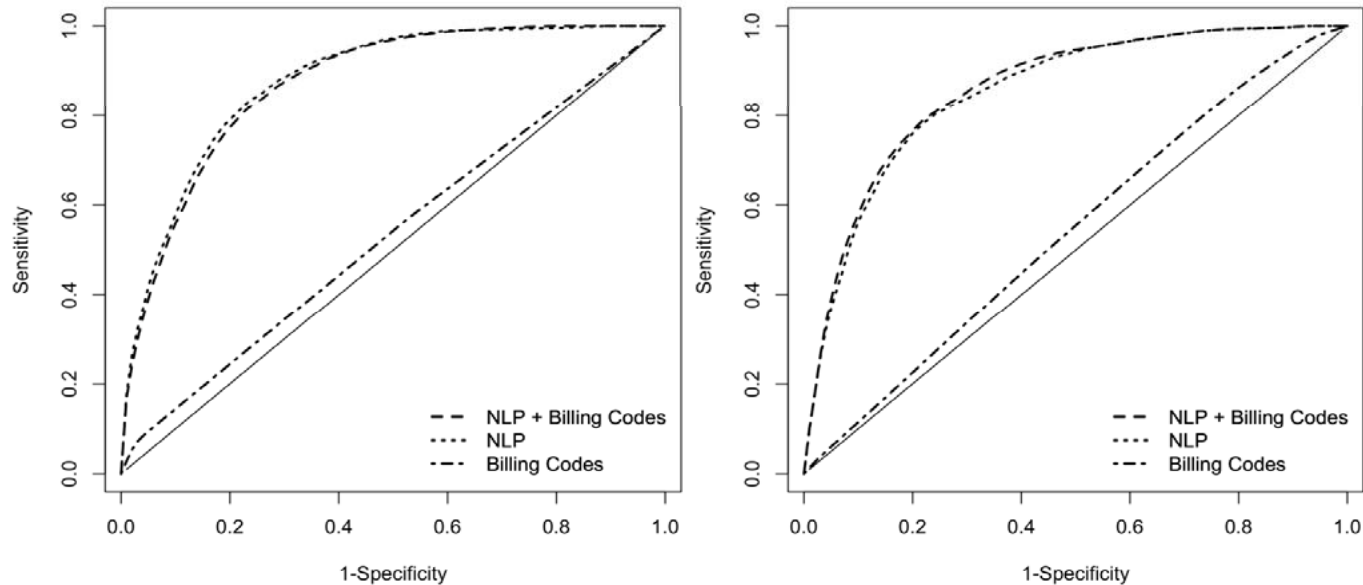
Cases - Positive Predictive Values



Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model

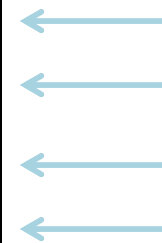
R. H. Perlis^{1,2*}, D. V. Iosifescu^{1,3}, V. M. Castro⁴, S. N. Murphy⁵, V. S. Gainer⁴, J. Minnier⁶, T. Cai⁶,
S. Goryachev⁴, Q. Zeng⁷, P. J. Gallagher², M. Fava¹, J. B. Weilburg¹, S. E. Churchill⁸,
I. S. Kohane⁹ and J. W. Smoller²

Use NLP to define cohorts of treatment-resistant and treatment-responsive depression

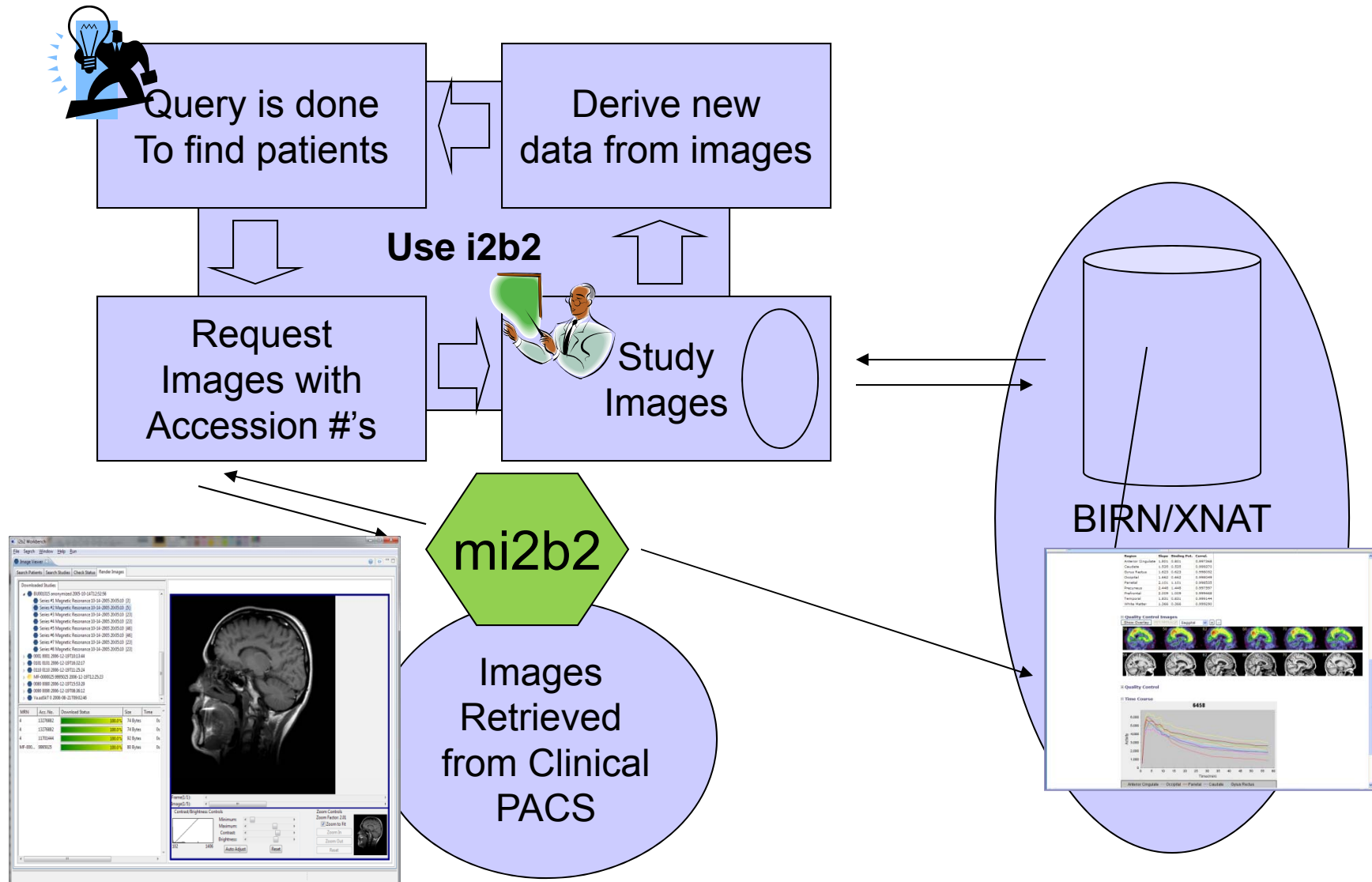


Specificity: 95%
AUC > 85%

Clinical Status	Model	Specificity	Sensitivity	Precision	AUC
Depressed	Billing Codes	0.95	0.09 (0.03)	0.57 (0.14)	0.54 (0.02)
Depressed	NLP	0.95	0.42 (0.05)	0.78 (0.02)	0.88 (0.02)
Depressed	NLP + Billing Codes	0.95	0.39 (0.06)	0.78 (0.02)	0.87 (0.02)
Well	Billing Codes	0.95	0.06 (0.02)	0.26 (0.27)	0.55 (0.03)
Well	NLP	0.95	0.37 (0.06)	0.86 (0.02)	0.85 (0.02)
Well	NLP + Billing Codes	0.95	0.39 (0.07)	0.85 (0.02)	0.86 (0.02)

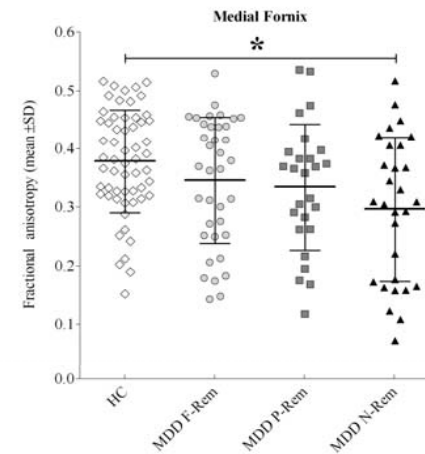
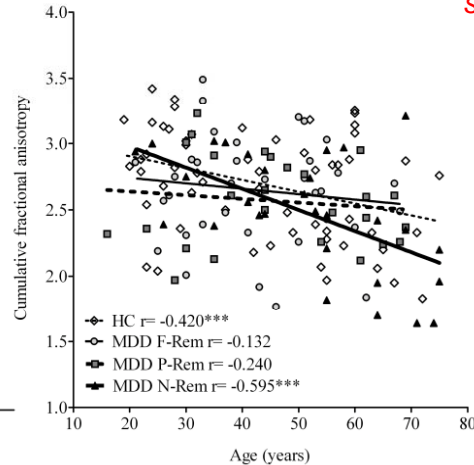
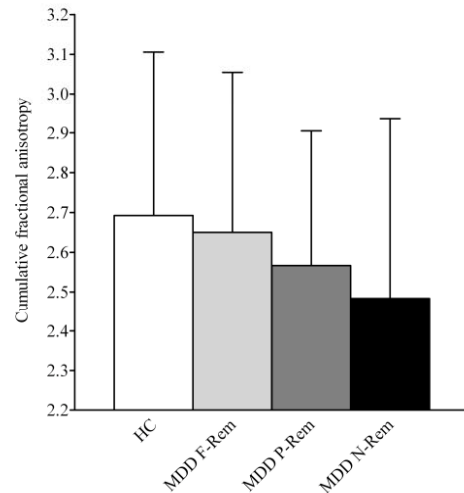
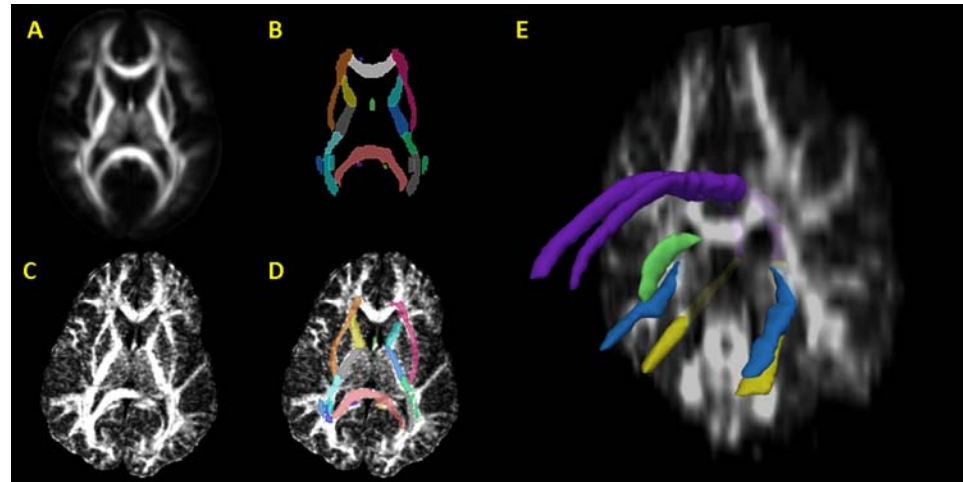


Medical Imaging Cell (mi2b2)



White matter abnormalities associated with treatment-resistant depression

- Scans collected as part of routine clinical care
- Diffusion tensor imaging in 150 pts
- Age-related decline in white matter integrity increases with treatment resistant depression



Rapid investigation of QTc prolongation

■ FDA warning 2011 for Celexa

Safety Announcement:

[8-24-2011] "should no longer be used at doses greater than 40 mg per day because it can cause abnormal changes in the electrical activity of the heart."

■ But, did NOT include Lexapro (which is active ingredient of Celexa [s-enantiomer])

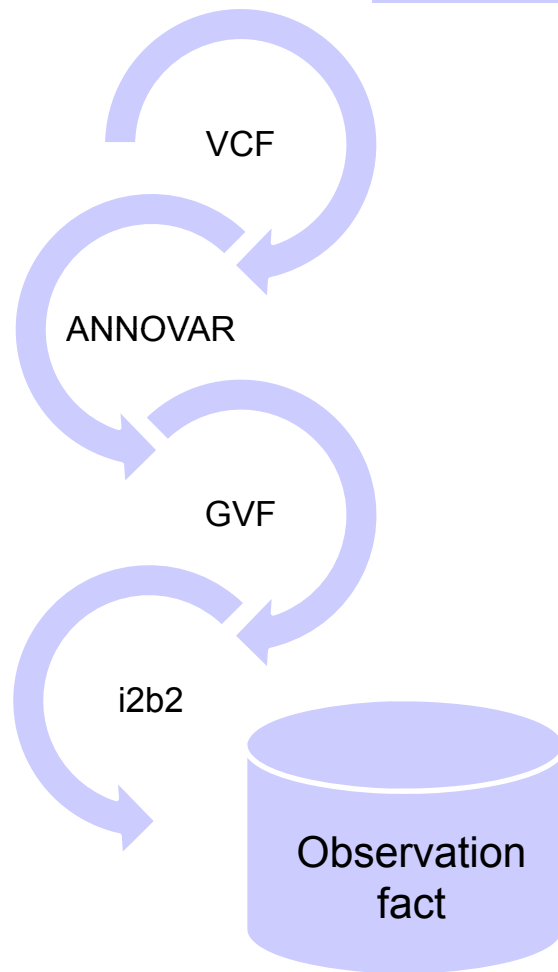
■ Shown to be true with RPDR-derived data set with >38,000 EKGs obtained within 14 – 90 day window after medication initiated

Anti-depressant	Adjusted model†	
	prolongatio n	p-value
SSRI		
Citalopram (Celexa)	2.85	0.004
Escitalopram (Lexapro)	3.80	< 0.001
Fluoxetine (Prozac)	1.44	0.150
Paroxetine (Paxil)	0.07	0.943
Sertraline (Zoloft)	0.87	0.383
Other anti-depressants		
Amitriptyline	4.10	< 0.001
Bupropion	-2.15	0.032
Duloxetine	0.60	0.547
Mirtazapine	-1.46	0.145
Nortriptyline	1.23	0.219
Venlafaxine	1.15	0.251
previously known prolonger		
Methadone	5.32	< 0.001

† Adjusted for age, gender, race, type of insurance, history of major depression, history of myocardial infarction and Charlson comorbidity score

Roy Perlis MD, MSc and team

Importing NGS variant output into i2b2



Variant Call Format

Gene Annotated VCF

Genome Variation Format

Pipeline - VCF to VCF-ANNO

1	1105366	.	T	C	.	PASS
	AA=T;AC=4;AN=114;DP=3251			GT:DP	1/0:54	

VCF

ANNOVAR

VCF-ANNO

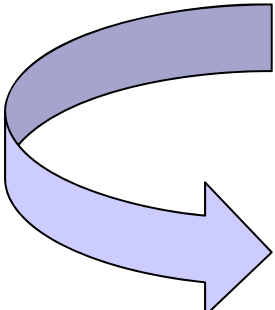
exonic	TLL10	1	1105366	1105366	T	C	1
	1105366	.	T	C	.	PASS	
	AA=T;AC=4;AN=114;DP=3251			GT:DP	1/0:54		

Pipeline - VCF-ANNO to GVF

exonic	TTL10	1	1105366	1105366	T	C	1
	1105366	.	T	C	.	PASS	
	AA=T;AC=4;AN=114;DP=3251		GT:DP		1/0:54		

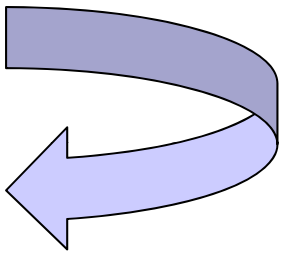
VCF-ANNO

VCF-ANNO
2GVF

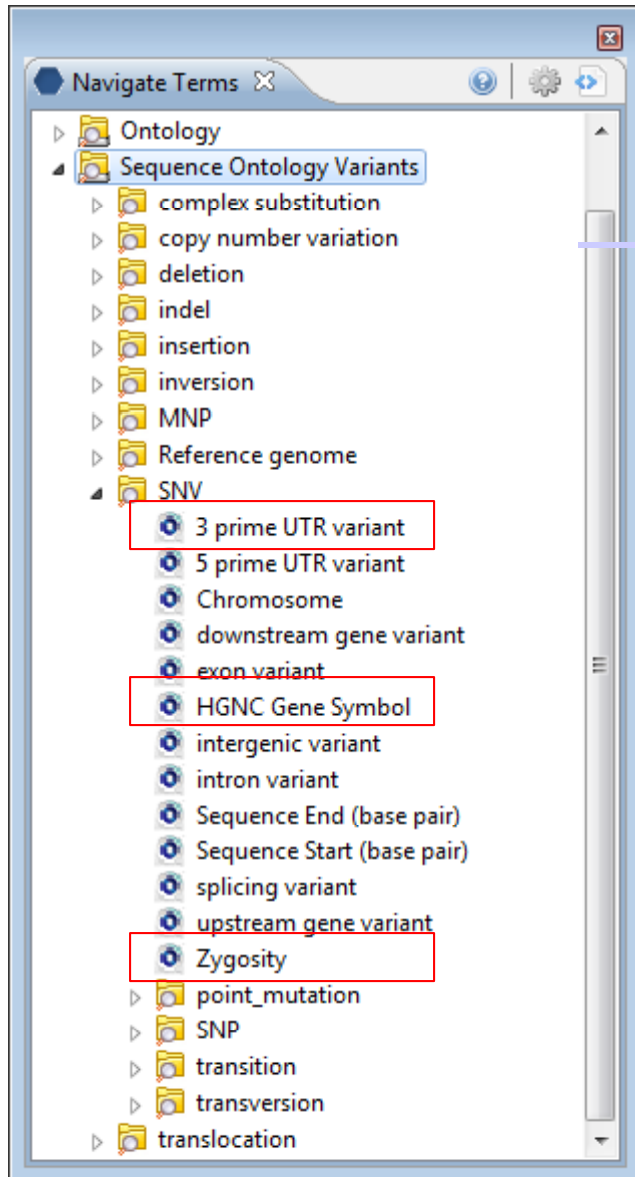


GVF

chr1	VCF	SNV	1105366	1105366	.	+
. ID=1;Reference_seq=T;Variant_seq=C;Variant_feature=exonic;Gene=TTL10; Genotype=heterozygous						



Navigating NGS Variant Data with Sequence Ontology



Combination of concepts and modifiers to identify:

An SNV located on a 3'UTR

An SNV associated with a certain gene

An SNV of specified zygotity

Querying NGS Variant Data

- Querying for a heterozygous SNV on an exon of gene TSHR
 - Note that all panels have same items instance

The screenshot shows the 'Query Tool' interface. At the top, there is a 'Query Name:' field. Below it are three query groups, each with its own 'Dates', 'Occurs > 0x', and 'Exclude' buttons, and a dropdown menu for 'Items instance will be same'. Group 1 contains the query: 'Gene Symbol [LIKE[contains] "TSHR."]'. Group 2 contains: 'SNV [exon variant]'. Group 3 contains: 'SNV [Zygosity Is heterozygous]'. To the right of the groups is an 'Analysis Types' panel with checkboxes for: Patient list, Event list, Number of patients (checked), Gender patient breakdown, Vital Status patient breakdown, Race patient breakdown, Age patient breakdown, and TimeLine (checked). Below this is a 'Query Timing' section with radio buttons for: Treat all groups independent, Selected groups occur in the, and Items instance will be same (selected). At the bottom, there is a 'Get Everyone' checkbox, a 'Run Query Above' button, and a 'Patient(s) returned:' field.

Gene Assist

The screenshot displays the i2b2 Workbench interface. The main window is titled "i2b2 Workbench" and shows a "Query Tool" tab. On the left, there is a "Navigate Terms" panel with a tree view of biological terms, including "SNV" and its sub-terms like "3 prime UTR variant", "5 prime UTR variant", "Chromosome", "downstream gene variant", "exon variant", "HGNC Gene Symbol", "intergenic variant", "intron variant", "Sequence End (base pair)", "Sequence Start (base pair)", and "splicing variant".

The "Query Tool" panel contains a "Query Name" field and three groups (Group 1, Group 2, Group 3) for defining query criteria. Each group has columns for "Dates", "Occurs > 0x", and "Exclude", with a "Treat Independently" dropdown. A term "SNV [HGNC Gene Symbol]" is added to Group 1. A "Choose modifier value of SNV" dialog box is open, allowing the user to search within the narrative text associated with the term "SNV". The dialog has three radio button options: "No Search Requested", "By abnormal flag", and "Search within Text" (which is selected). There is also a "Contains..." search field and "OK", "Cancel", and "Gene Assist" buttons.

On the right side, there are two panels: "Analysis Types" and "Query Timing". The "Analysis Types" panel includes checkboxes for "Patient list", "Event list", "Number of patients" (checked), "Gender patient breakdown", "Vital Status patient breakdown", "Race patient breakdown", "Age patient breakdown", and "TimeLine". The "Query Timing" panel has radio buttons for "Treat all groups independently" (selected), "Selected groups occur in the same instance", and "Items instance will be same".

At the bottom, there are two panels: "Analysis details" and "Sample details". The "Analysis details" panel includes fields for "Specify input file:", "Input file format:" (set to "VCF"), "VCF mapping file:", "I2B2 Patient number:", "I2B2 Encounter number:", "Date of encounter:" (with a "YYYY-MM-DD HH:MM:SS format" hint), and "Reference genome version:" (set to "hg18"). The "Sample details" panel includes fields for "Sample ID:", "Sample Type:" (set to "TISSUE"), "Anatomical Source:" (set to "Pericardium"), "Collection Method:" (set to "BIOPSY"), "Additive:" (set to "UNKNOWN"), and "Sample Pathology" (set to "TUMOR").

Gene Assist

The screenshot displays the i2b2 Workbench interface. The main window is titled "i2b2 Workbench for Demo" and includes a "Query Tool" and "Analysis View" section. The "Query Tool" shows a "Query Name" field and three groups of filters (Group 1, Group 2, Group 3) with columns for "Dates", "Occurs > 0x", and "Exclude". The "Analysis View" shows "Analysis Types" with options like "Patient list", "Event list", "Number of patients", "Gender patient breakdo", and "Vital Status patient bre".

A "Genomic Browser" window is overlaid on the main interface, displaying the HGNC (HUGO Gene Nomenclature Committee) Quick Gene Search page. The page features a search bar, navigation links (Home, Search Genes, Downloads, Gene Families, HCOP, Useful Links, About, Contact Us), and a "Request Symbol" button. The search criteria are set to "Search symbols, keywords or IDs for:" with options for "equal", "begin", and "contain", and a "Display 50 hits" option. The footer of the HGNC page includes logos for EMBL-EBI, NHGRI, and Wellcome Trust, along with contact information and a link to the Admin page.

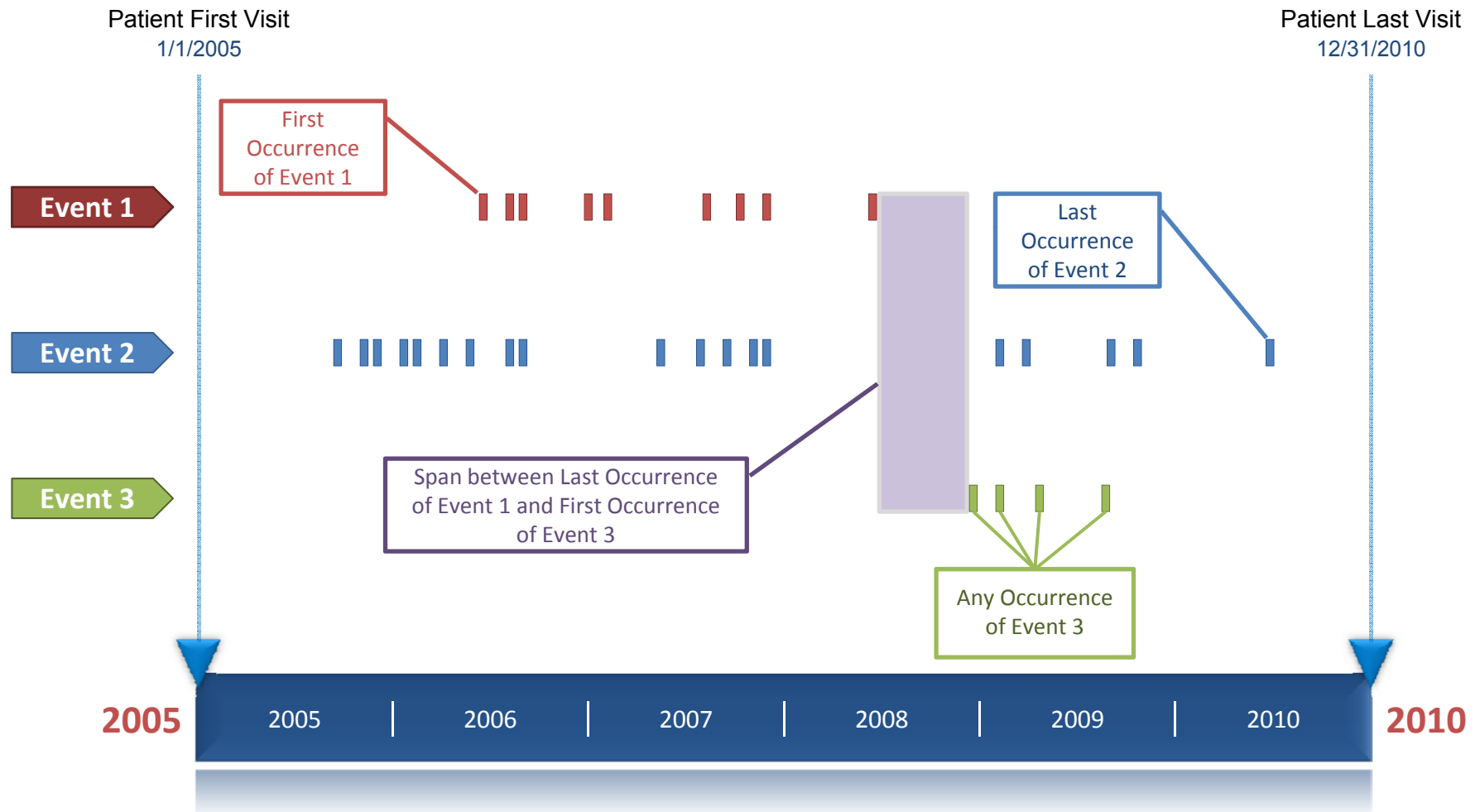
The "Previous Queries" section on the left lists various queries such as "Chromos-exon va@09:19:34 [03-08-2013] [lcp5]", "Chrom-exon -Diabe@09:12:00 [03-08-2013] [lcp5]", "Chrom-exon -Diabe@09:05:59 [03-08-2013] [lcp5]", "Chromos-exon va@09:04:52 [03-08-2013] [lcp5]", "Chromosome@09:01:49 [03-08-2013] [lcp5]", "interge-Chromos@07:28:38 [03-08-2013] [lcp5]", "HGNC -exon -Zygos@18:10:27 [03-07-2013] [lcp5]", "EGFR-exon-heterzygous [03-05-2013] [lcp5]", "insertion@10:08:54 [03-04-2013] [lcp5]", "deletion@09:56:53 [03-04-2013] [lcp5]", "EN2-AUT@04:09:55 [02-28-2013] [lcp5]", "TSHR-hyperthyroidism@04:02:27 [02-28-2013] [lcp5]", "substit-substit@03:41:46 [02-28-2013] [lcp5]", "MAD1L1-prostateCancer@03:34:56 [02-28-2013] [lcp5]", "substitution@03:31:21 [02-28-2013] [lcp5]", "subst-subst-subst@03:27:17 [02-28-2013] [lcp5]", and "substitution@09:26:27 [02-27-2013] [lcp5]".

The "Analysis details" section on the right includes fields for "Date of encounter:" (YYYY-MM-DD HH:MM:SS format), "Reference genome version:" (hg18), and "Sample Pathology" (TUMOR).

New Core i2b2 Features

- Top Level 1.7 features
 - Temporal Queries Enabled
 - Identity Management cell to manage patient participation in i2b2 projects
 - Patient management for Out-of-the-box clinical trials support
 - EMR views of patient for translational medicine and clinical trials support

Temporal Query Formulation



Temporal Query Definition

- Temporal queries are organized around the concept of an event.
- Events are related to each other through “subquery constraints”
 - An event is defined from an observation
 - Start date or end date
 - First, last, *Any*
 - Events are compared in time to be
 - Equal, Less, Greater, Less Than Equal, Greater Than Equal
 - Within a certain time span
 - Year, Month, Day, Hr, Min
- Each event is defined as a “subquery” within the query definition
 - In the XML is a Recursive definition – a subquery is a query

Defining Temporal Relationships

The screenshot displays the i2b2 Workbench interface for i2b2 Demo 1.7 on Oracle. The main window is titled "Query Tool (Experimental)" and contains the following components:

- Query Name:** A text input field.
- Specify Temporal Query:** A dropdown menu.
- Events in Query:** A list of events with their constraints:
 - Event 1:** No Date Constraints
 - Group 1:** Observation, No Date Constraints, >0 (Asthma)
 - Event 2:** No Date Constraints
 - Group 1:** Observation, No Date Constraints, >0 (Albuterol)
- Define Temporal Relationships Among Events:** A configuration panel for temporal relationships:
 - Start of:** the First Ever, Event 1
 - Occurs Before:** dropdown menu
 - Start of:** the First Ever, Event 2
 - By:** ≤, 6, month(s)
 - And:** >, 1, day(s)
 - Add Temporal Relationship:** button
- Event Inclusion:** A section with a checked checkbox for "Only query for Events used in Temporal Relationships".
- Submit Query:** button
- Patient(s) Returned:** 6

On the left side, there are two panels:

- Navigate Terms:** A tree view showing a hierarchy of terms including Route, Rx Sig, Adrenergic bronchodilator, Patient Instructions, Route, Rx Sig, Albuterol, Bitolterol, Epinephrine, Formoterol, and Isoetharine.
- Previous Queries:** A search interface with a "Search By Name" dropdown (set to "Containing"), a "Find" button, and a "Any Category" dropdown. Below this is a list of previous queries with timestamps, such as "Test Temporal Query @3/19/2013 11:29:2".

Viewing Temporal Relationships

The screenshot displays the i2b2 Workbench interface for i2b2 Demo 1.7 on Oracle. The main window is titled "i2b2 Workbench for i2b2 Demo 1.7 on Oracle" and shows the user "i2b2 User" with a status of "i2b2".

The interface is divided into several panels:

- Navigate Terms:** A tree view on the left showing various medical terms such as "Patient Instructions", "Route", "Rx Sig", "Antiasthmatic combinations", "Antihistamines", "Antitussives", "Bronchodilators", "Adrenergic bronchodilators", "Albuterol", "Bitolterol", "Epinephrine", and "Formoterol".
- Query Tool (Experimental):** A central panel for configuring queries. It includes a "Query Name" field, a "Specify Temporal Query" dropdown, and a table for "Events in Query". The table lists "Event 1" and "Event 2", both with "No Date Constraints". Below the table is an "Add Event" button. A "Define Temporal Relationships Among Events" dialog is open, showing configurations for "Event 1" and "Event 2". The dialog includes dropdowns for "Start of", "the First Ever", and "Event 1/2", a "Occurs Before" dropdown, and a "By" field set to "6 month(s)". There is also an "And" field set to "1 day(s)". An "Add Temporal Relationship" button is at the bottom.
- Event Inclusion:** A section with a checked option "Only query for Events used in Temporal Relationships".
- Submit Query:** A button that has been clicked, resulting in "Patient(s) Returned: 6".
- Align-in-time View:** A panel at the bottom showing a timeline view. It has tabs for "Record View", "Comparison View", and "Group View". The "Record View" is active, showing "All Records" (Records 6/6). The timeline is zoomed to "Years" and annotated. The records are listed on the left, including "Albuterol" and "Asthma" for various patient IDs. The timeline shows green triangles for "Albuterol" and orange triangles for "Asthma" events.
- Previous Queries:** A panel on the left with a search bar and a list of previous queries, including "Test Temporal Query @3/19/2013 11:29:28.240" and "Diabete-Metform@10:52:25 [03-19-2013] [demo]".

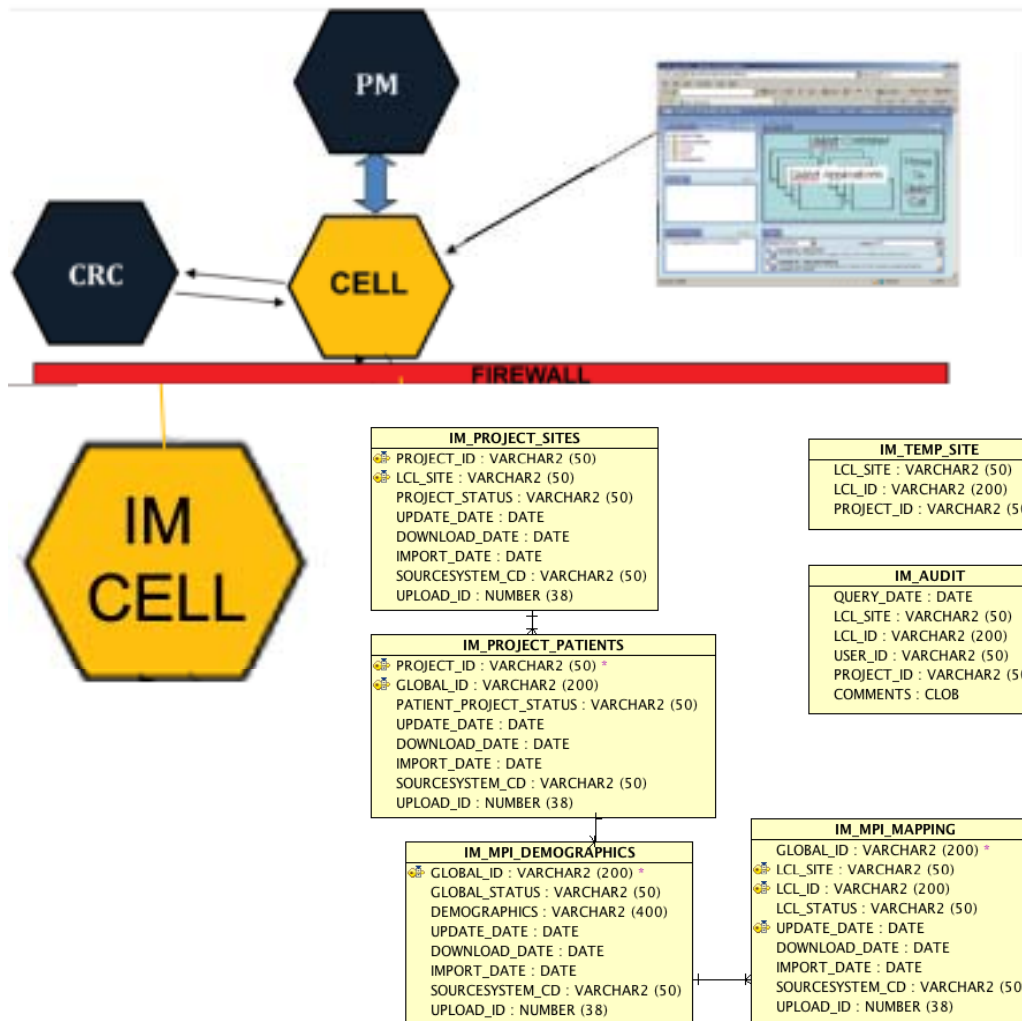
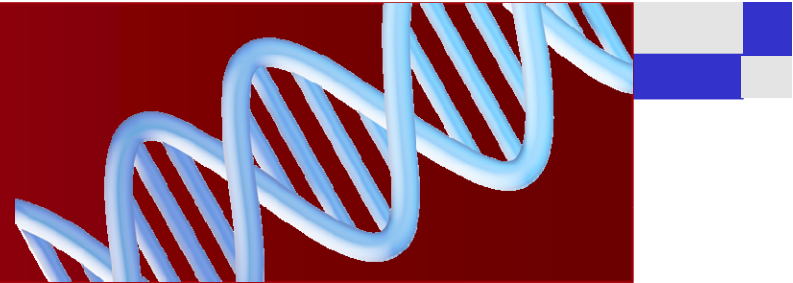
Patient can be conceived throughout i2b2 as single item

The screenshot displays the i2b2 Workbench interface. The main window is titled "i2b2 Workbench for i2b2demodata2" and includes a menu bar (File, Search, Window, Help, Run) and a status bar (demo Status: i2b2). The interface is divided into several panes:

- Navigate Terms:** A tree view on the left showing categories like Clinical Trials, Custom Metadata, Demographics, Diagnoses, Expression Profiles Data, ICD-10CM (NCBO), Laboratory Tests, Medications, Procedures, Providers, Reports, and Visit Details.
- Query Tool:** The central area for building queries. It features three groups (Group 1, Group 2, Group 3) for organizing terms. Each group has options for "Dates", "Occurs > 0x", and "Exclude". The "Treat Independently" dropdown is selected for each group. A "PATIENT HIVE: 100000011" is listed in Group 1. Below the groups are instructions: "Drag terms from Navigate, Find and Workplace into this group".
- Workplace:** A tree view on the right showing a "demo" folder containing "Study Patients" and "PATIENT HIVE:100000011".
- Analysis Types:** A dropdown menu on the right with options: "Patient set", "Encounter set", and "Number of patients".
- Query Timing:** A section for monitoring query performance.
- Buttons:** "Get Everyone", "Run Query Above", and "Patient(s) returned:".
- Previous Queries:** A section for managing saved queries.
- Patient Set Viewer:** A section for viewing patient sets. It includes a search field and a list of patient sets. The selected set is "Patient Set for '(PrevQu+Hyperte@01:37:41)'".
- Table:** A table displaying patient data for the selected set. The columns are HIVE, EMP, BWI, MGH, NWH, BWH, NSMC_SALEM, NSMC_UNION, FH, and DFC.
- Bottom Bar:** "Retrieving information on this item ...".

HIVE	EMP	BWI	MGH	NWH	BWH	NSMC_SALEM	NSMC_UNION	FH	DFC
100000011	100790915	01164887	3000001831	4000002011	11489929	S500003061	U500004021		
100000025	101809330	01800290	3000001845	4000002025	15783376	S500003075	U500004035		
100000026	102344360	01954309		4000002026	17028580	S500003076	U500004036	30001003	252304
100000043	102344370	01954388	3000001853	4000002043	17028663		U500004053	30001005	252311
100000045	102344373	01954389	3000001855	4000002045	17028671	S500003095	U500004055		252312
100000054	102637795	01722840	3000001874	4000002054	18092957	S500003104			
100000071	102788263	02161640	3000001891	4000002071	18658583	S500003121	U500004081		
100000083	103703039	02408012	3000001903	4000002083	20722294		U500004093		346657
100000087	104308528	01517492	3000001907		00000091	S500003137	U500004097		
100000089	104334898	01164890	3000001909	4000002089	11489952	S500003139	U500004099		299566
100000096	105511313		4745320	4000002096	2000002056	S500003116	U500004106		
100000105	105560546		4745320	4000002105		S500003155	U500004115		
100000108	105560549		4745322	4000002108	2000002068	S500003158	U500004118		

I2B2 IM ARCHITECTURE



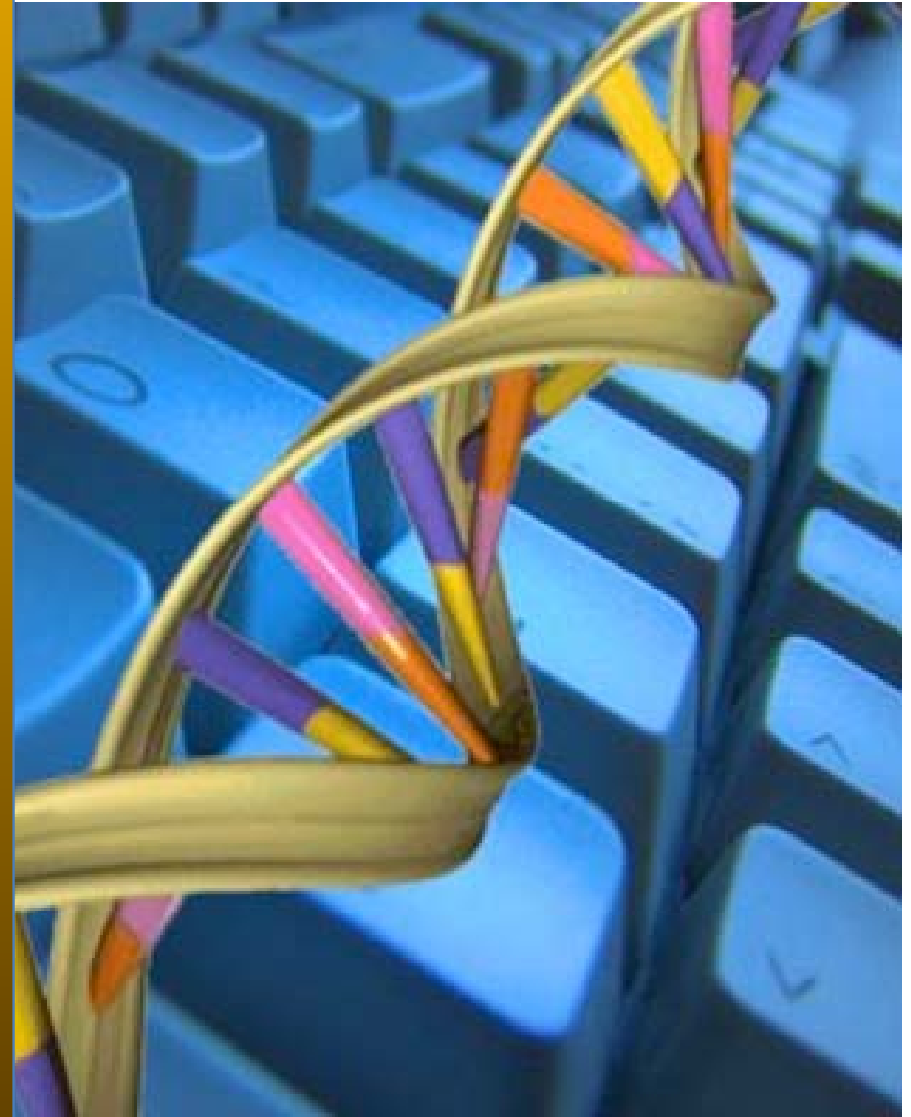
- Identity Management (IM) cell becomes a primary core cells within the i2b2 hive
- Capable of converting i2b2 patient numbers back into identifying MRNs
- Retrieving and storing Protected Health Information (PHI)
- Critical for investigators who plan to run a study and recruit patients
 - Lists of patients with real identifiers are managed and linked to a project
- Allows multiple identifiers to be resolved by an Enterprise Master Patient Indexes

JBoss 7

- ✓ Improved Performance
 - ✓ Better Security
 - ✓ J2EE 6
- ✓ Administration Improvements
 - ✓ Integrate into Eclipse



PARTNERS™
HEALTHCARE SYSTEM

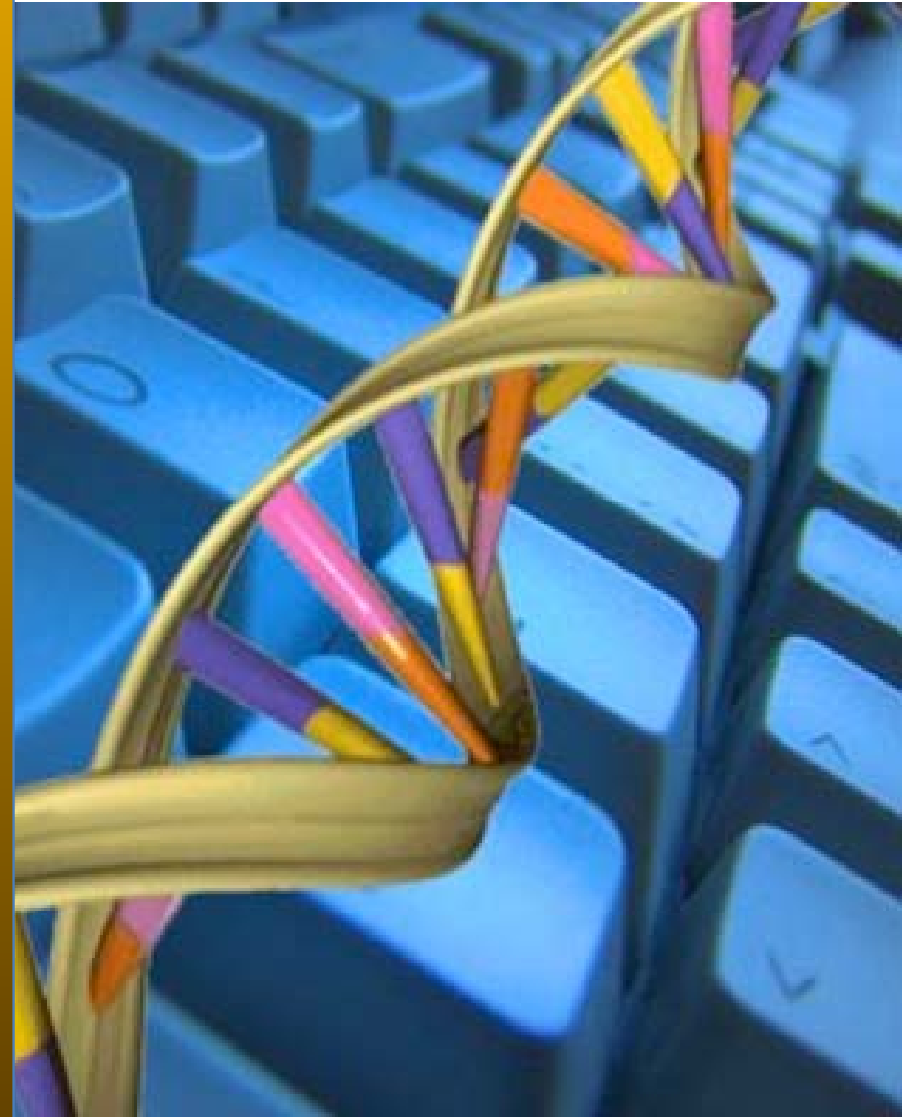


IDENTITY MANAGEMENT

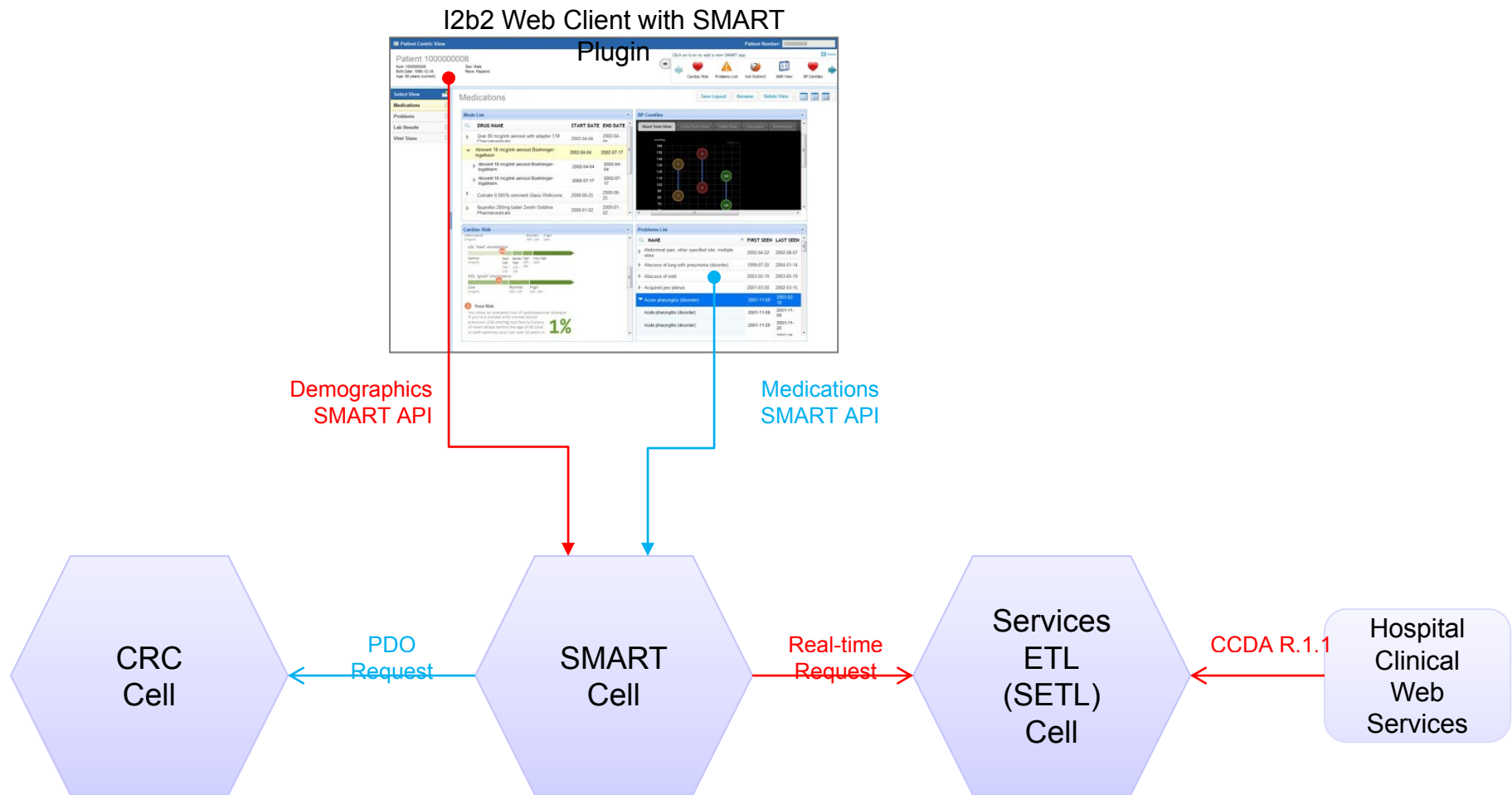
- ✓ Core Cell
- ✓ Encrpytion
- ✓ Auditing
- ✓ New for 1.7



PARTNERS™
HEALTHCARE SYSTEM



Development of new Services ETL (SETL) Cell



Services ETL: Output PDO Response

```
<ns2:patient_set>
  <patient>
    <patient_id source="hive">XXXXXXXXXX</patient_id>
    <param column="vital_status_cd" name="vital_status_cd">U</param>
    <param column="birth_date" name="birth_date">19490101</param>
    <param column="sex_cd" name="sex_cd">M</param>
    <param column="language_cd" name="language_cd">SPANISH</param>
    <param column="religion_cd" name="religion_cd">PROTESTANT</param>
    <param column="race_cd" name="race_cd">OTHER</param>
    <param column="ethnicity_cd" name="ethnicity_cd">AFRICAN</param>
    <param column="marital_status_cd" name="marital_status_cd">UNKNOWN</param>
    <param column="legal_first_name" name="legal_first_name">SANTA</param>
    <param column="legal_middle_initial" name="legal_middle_initial">J</param>
    <param column="legal_last_name" name="legal_last_name">CLAUS</param>
    <param column="legal_suffix" name="legal_suffix">JR</param>
    <param column="permanent_line1" name="permanent_line1">1010 TENTH
      STREET</param>
    <param column="permanent_line2" name="permanent_line2">APT 110</param>
    <param column="permanent_city" name="permanent_city">BOSTON</param>
    <param column="permanent_state" name="permanent_state">MA</param>
    <param column="permanent_zip" name="permanent_zip">02114</param>
    <param column="permanent_country" name="permanent_country">US</param>
    <param column="local_line1" name="local_line1">55 FRUIT ST</param>
    <param column="local_line2" name="local_line2">APT 2</param>
    <param column="local_city" name="local_city">BOSTON</param>
    <param column="local_state" name="local_state">MA</param>
    <param column="local_zip" name="local_zip">02114</param>
    <param column="local_country" name="local_country">US</param>
    <param column="primary_phone" name="primary_phone">9781231231217</param>
    <param column="work_phone" name="work_phone">78144455553333</param>
    <param column="mobile_phone" name="mobile_phone">6039275569</param>
    <param column="other_phone" name="other_phone">(987) 111-1111</param>
  </patient>
</ns2:patient_set>
```

Launching 'standalone' SMART-i2b2 views

The screenshot displays the i2b2 Workbench interface. The main window shows a patient-centric view for Santa J Claus, with a callout box stating: "Demographics data is retrieved in real-time from the SETL cell". The interface includes a navigation pane on the left, a query tool at the top, and a main content area with a patient summary, a medications list, and a problems list. A blue arrow points to a specific patient entry in the timeline view.

Demographics data is retrieved in real-time from the SETL cell

Medications List

DRUG NAME	START DATE	END DATE
Qvar 80 mcg/inh aerosol with adapter 3 M Pharmaceuticals	2002-04-04	2002-04-04
208 ACTUAT Ipratropium Bromide 0.018 MG/ACTUAT Metered Dose Inhaler [Atrivent]	2002-04-04	2002-07-17
Cutivate 0.005% ointment Glaxo Wellcome	2008-06-25	2008-06-25
Ibuprofen 200mg tablet Zenith Goldline Pharmaceuticals	2009-01-02	2009-01-02
Cyclobenzaprine Hydrochloride 10mg tablet Geneva Pharmaceuticals	2009-01-22	2009-01-22
nabumetone 750 MG Oral Tablet [Relafen]	2009-01-28	2009-01-28
Acetaminophen-Codeine Phosphate		

Problems List

NAME	START DATE	END DATE
Abdominal pain, other specified site, multiple sites	2002-08-07	2002-08-07
Abdominal pain, other specified site, multiple sites	2002-04-29	2002-04-29
Abdominal pain, other specified site, multiple sites	2002-04-22	2002-04-22
Abscess of lung with pneumonia (disorder)	2004-01-14	2004-01-14
Abscess of lung with pneumonia (disorder)	2004-01-02	2004-01-02
Abscess of lung with pneumonia (disorder)	2002-11-14	2002-11-14
Abscess of lung with pneumonia (disorder)	2002-08-06	2002-08-06
Abscess of lung with pneumonia (disorder)	1999-10-26	1999-10-26
Abscess of lung with pneumonia (disorder)	1999-10-25	1999-10-25
Abscess of lung with pneumonia (disorder)	1998-07-28	1998-07-28
Abscess of lung with pneumonia (disorder)	1998-07-26	1998-07-26
Abscess of lung with pneumonia (disorder)	1999-07-26	1999-07-26
Abscess of lung with pneumonia (disorder)	1999-07-20	1999-07-20

Timeline View

- Person_#10000
- CPK_(Group:CPK)
- Person_#10000
- CPK_(Group:CPK)
- Person_#100000025_Female_76yroid_Hispanic
- CPK_(Group:CPK)
- Person_#100000027_Male_37yroid_Unknown
- CPK_(Group:CPK)
- Person_#100000055_Female_19yroid_Hispanic
- CPK_(Group:CPK)

Patient Set: Patient Set: 43 Patients start: 11 increment: 10

Supports workflow for Clinical Trials

- Person does query as obfuscated user in large data mart
- Optimal query results can be used to create request for approval so that patients can be viewed as a limited data set
- Approval is obtained and a new project is created where those patients in the Optimal patient set can be viewed in plug-ins such as timeline, charts, and de-identified SMART views.
- Patients are chosen that represent truly Optimal patients for the Clinical trial.
- PHI is viewed on the truly Optimal patients in a specially Audited view that may be governed by the local application of “special rules”

Workflow to support Clinical Trials

The screenshot displays the i2b2 Web Client interface in Mozilla Firefox. The browser address bar shows `webservices.i2b2.org/webclient/`. The page title is "i2b2 Query & Analysis Tool". The user is logged in as "i2b2 User" for "Project i2b2 Demo".

The interface is divided into several sections:

- Navigate Terms:** A tree view on the left containing categories like Clinical Trials, Custom Metadata, Demographics, Diagnoses, Expression Profiles Data, Laboratory Tests, Medications, and Procedures.
- Workplace:** A file explorer showing a "demo" folder with subfolders for "All Patients" and "Approved Patients".
- Project Request:** A table with columns for Patient ID and Project Variables (SZ, CVD, HIV, DM). The table contains five rows of patient data.
- Previous Queries:** A list of recent queries with timestamps and user information.
- Plugins:** A section at the bottom right showing "Detailed List View" and "Category: ALL", along with "Timeline" and "Project Request" plugin descriptions.

Two red annotations with arrows point to specific elements in the Project Request table:

- Drag and Organize Patients through Workplace:** An arrow points from the "Workplace" section to the "Approved Patients" folder.
- Deep dive into Optimal Patients for clinical trial:** An arrow points to the row for Patient ID 59, which has the text "67y M" in blue.

Patient ID	Project Variables					
		SZ	CVD	HIV	DM	
37	19y M	Y	N	N	Y	
38	37y F	N	Y	N	Y	
59	67y M	Y	N	Y	Y	
60	56y M	Y	Y	Y	Y	
71	22y F	Y	N	Y	Y	

Workflow to support Clinical Trials

- EMR View that looks familiar to clinicians
- Patients can be screened using SMART App's

The screenshot displays the i2b2 Web Client interface. The browser address bar shows 'smarti2b2/webclient'. The main header is 'i2b2 Query & Analysis Tool' with navigation links: 'Find Patients | View Patient | Analysis Tools | Message Log | Help | Logout'. The patient ID is '1000000059 [35 y/o M hispanic]'. The 'SMART EMR View' section shows patient details: 'Patient 1000000059', 'Num: 1000000059', 'Sex: Male', 'Born: 1975-02-21 (Age: 37)', 'Race: Hispanic', 'Language: German', and 'Zip Code: 02458'. A 'Select View' sidebar lists 'Medications', 'Problems', 'Allergies', 'Labs', and 'Notes'. The 'Problems' section is active, showing a table of medical issues. A 'Cardiac Risk' widget displays cholesterol levels and a risk assessment of 2%.

NAME	FIRST SEEN	LAST SEEN
Accidents caused by other powered hand tools	2005-01-18	2005-01-19
Acute contact otitis externa	2001-05-16	2001-05-16
Acute pharyngitis (disorder)	2005-08-05	2005-08-05
Acute upper respiratory infection (disorder)	2004-05-27	2004-05-27
Allergic rhinitis (disorder)	2006-06-30	2006-06-30
Asthma, unspecified without mention of status asthmaticus	1997-11-04	2008-11-26
Benign neoplasm of connective/soft tissue of	1998-04-09	1998-04-09

Cardiac Risk

LDL "bad" cholesterol: 146 mg/dL. Optimal: 0 mg/dL. Near Opt: 100-129. Border, High: 129-159. High: 160-190. Very High: 190+.

HDL "good" cholesterol: 34 mg/dL. Low: 0 mg/dL. Normal: 100-129. High: 129-159.

3 Your Risk

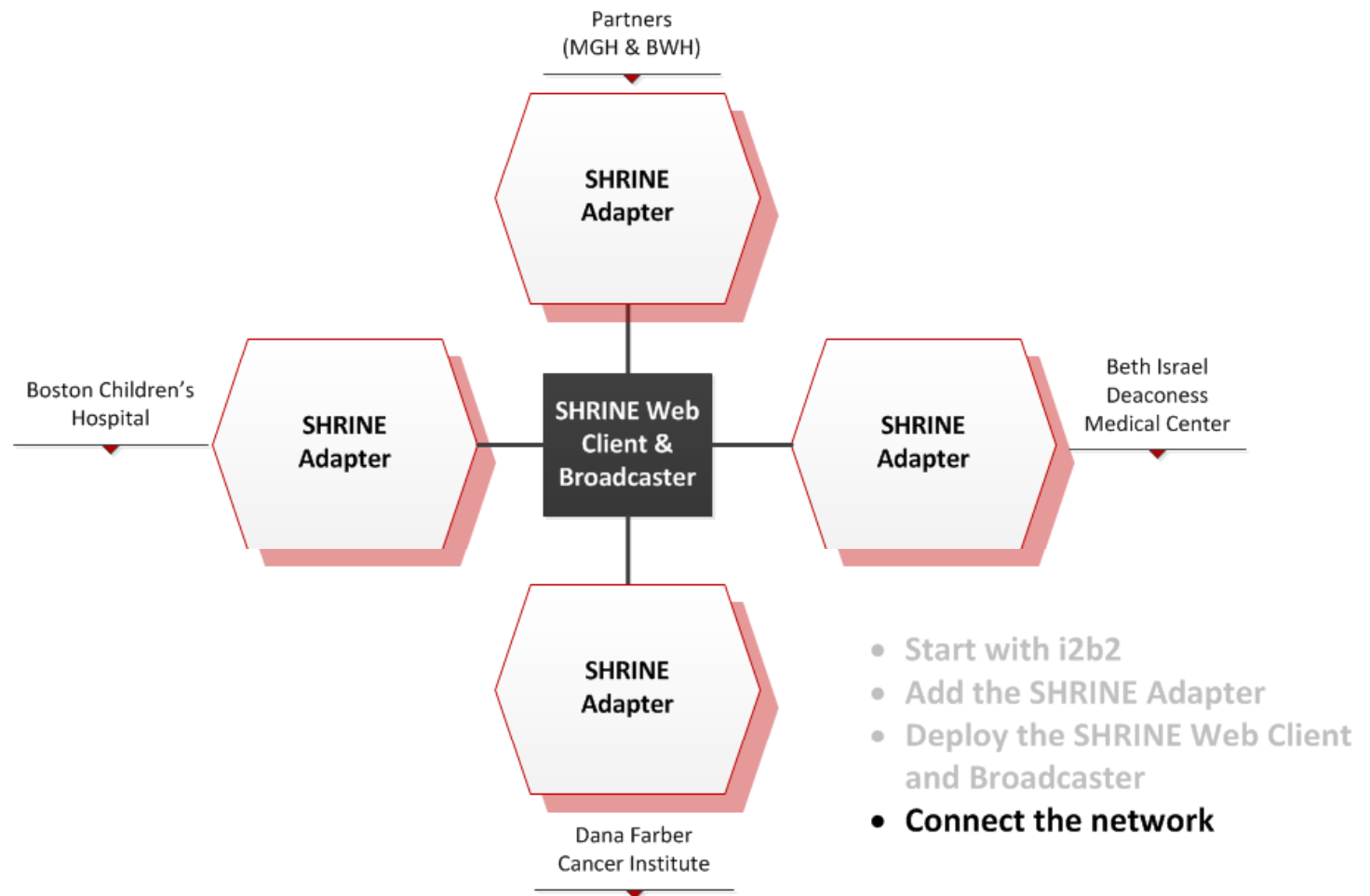
You show an elevated risk of cardiovascular disease. If you're a smoker with normal blood pressure, (130 mm/Hg) but family history of heart attack before the age of 60 (one or both parents) your risk over 10 years is: **2%**



Timelines for new i2b2 Releases

- Release Candidate of Version 1.7 Core in April
- Clinical trials out-of-the-box plug-ins in September
- Next-Generation Sequencing plug-ins in September

Harvard SHRINE Architecture



Harvard SHRINE Metrics

- **6M patients**
- **1B facts**
- **4 ontology categories, 18K terms**
- **10K potential users**
- **5 IRBs**
- **5 major competing hospitals, 4 sites**
 - Partners HealthCare (BWH, MGH)
 - Boston Children's Hospital
 - Beth Israel Deaconess Medical Center
 - Dana-Farber Cancer Institute



SHRINE Find Patients | Message Log | Help | Logout

Navigate Terms Find Terms

- SHRINE
 - Demographics
 - Diagnoses
 - Certain conditions originating in the perinatal period
 - Complications of pregnancy, childbirth, and the puerperium
 - Congenital anomalies
 - Diseases of the blood and blood-forming organs
 - Diseases of the circulatory system
 - Diseases of the digestive system
 - Diseases of the genitourinary system
 - Diseases of the musculoskeletal system and connective tissue
 - Diseases of the nervous system and sense organs
 - Diseases of the respiratory system
 - Diseases of the skin and subcutaneous tissue
 - Endocrine, nutritional, and metabolic diseases and immunity disorders
 - Infectious and parasitic diseases
 - Injury and poisoning
 - Mental illness
 - Adjustment disorders
 - Alcohol-related disorders
 - Anxiety disorders
 - Attention deficit, conduct, and disruptive behavior disorders
 - Delirium, dementia, and amnesic and other cognitive disorders

Query Tool

Query Name: Pervasi-0-9 yea@00:44:10

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
				0-9 years old				
				10-17 years old				
				18-34 years old				

one or more of these
AND
one or more of these
AND
drag a term to here

Autism Info Request New Topic

Run Query New Query Print Query 2 Groups New Group

Previous Queries

- Pervasi-0-9 yea@00:44:10 [9-27-2010] [kohane]
- Perva-0-9 y-PHENO@17:57:23 [9-26-2010] [kohane]
- PDD-0-34@17:40:42 [9-26-2010] [kohane]
- Perva-Male-Schiz@16:27:52 [9-26-2010] [kohane]
- AI+PDD-0-34@17:47:17 [9-26-2010] [kohane]
- medicated-ppd-34@16:41:28 [9-26-2010] [kohane]
- =34@16:39:02 [9-26-2010] [kohane]
- Pervasi-0-9 yea@16:37:13 [9-26-2010] [kohane]

Query Status

Finished Query: "Pervasi-0-9 yea@00:44:10"

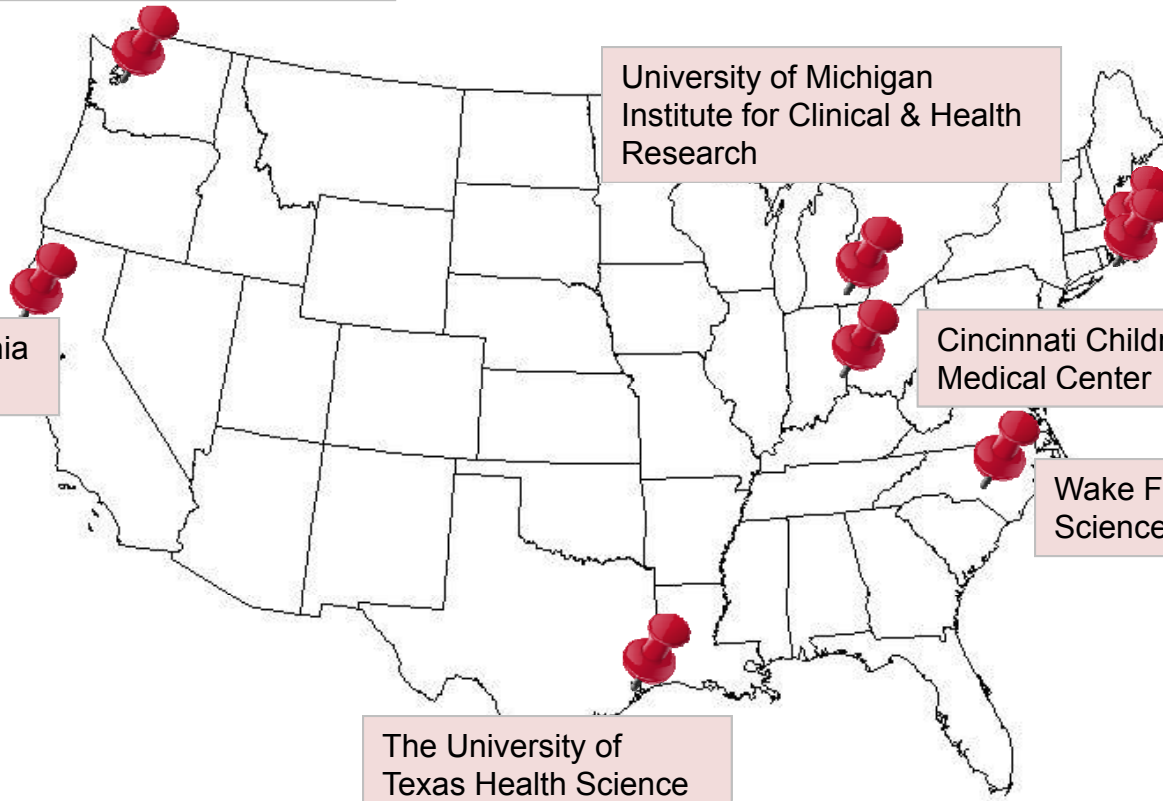
BIDMC - 141 ±3 patients	FINISHED [78.7 secs]
CHB - 9103 ±3 patients	FINISHED [78.7 secs]
Partners - 5134 ±3 patients	FINISHED [78.7 secs]

National SHRINE Demo

- **A demonstration of the feasibility of a national research network**
- **8 selected sites to perform queries on co-morbidities related to the primary diagnoses of autism and diabetes**
 - Patient must be diagnosed with either Autism Spectrum Disorders or Diabetes Mellitus at least once between 6/01/09 – 5/31/11

8 Participating Institutions

The Institute of Translational Health Sciences,
University of Washington



University of Michigan
Institute for Clinical & Health
Research

Partners HealthCare

Boston Children's
Hospital

University of California
San Francisco

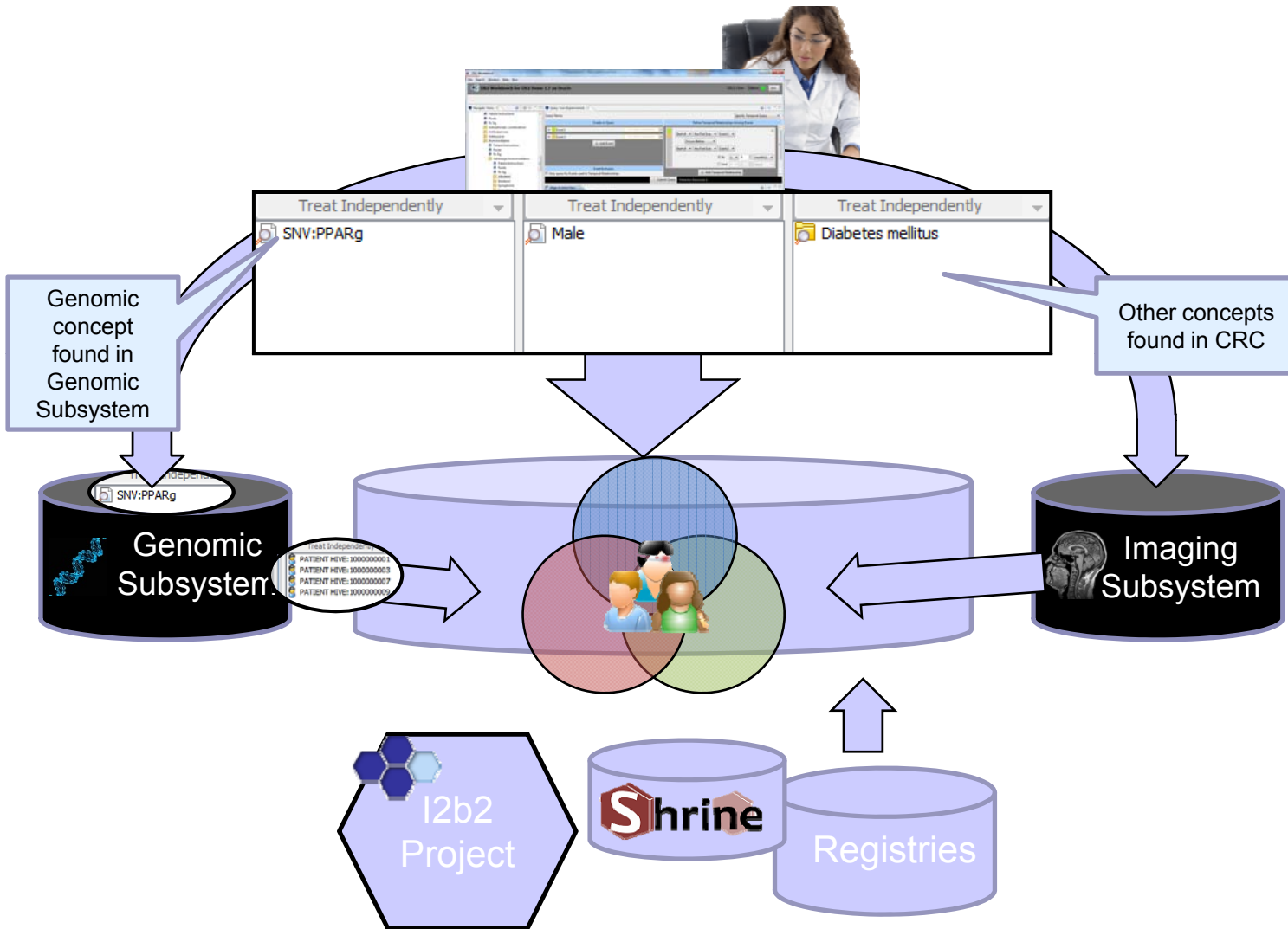
Cincinnati Children's Hospital
Medical Center

Wake Forest University Health
Sciences

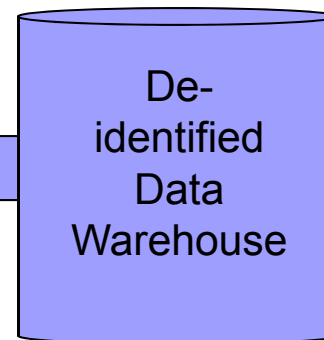
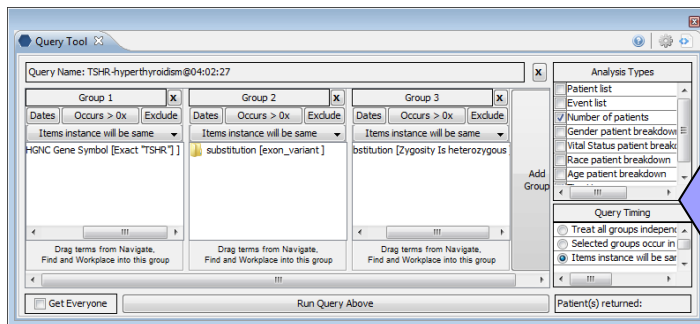
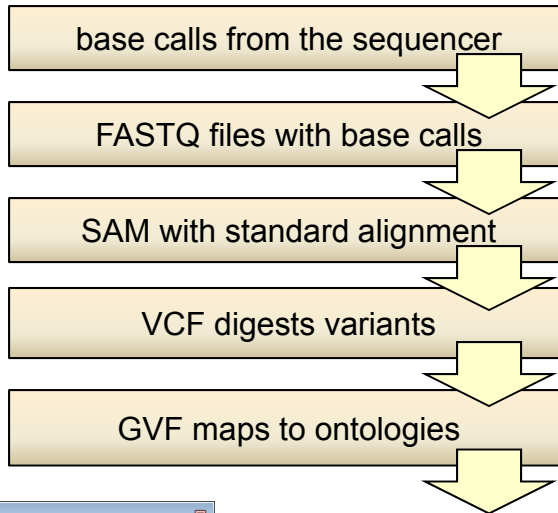
The University of
Texas Health Science
Center at Houston

- **Next major version of SHRINE – focused on supporting clinical trials**
- **Start with SHRINE aggregate query**
- **Narrow/validate set of subjects by incrementally filtering on detailed patient data**
- **Establish a set of subjects for recruitment into a clinical trial**

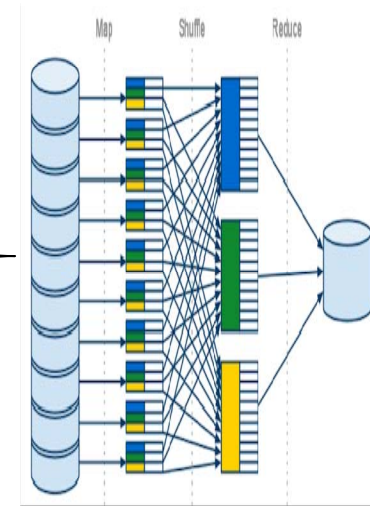
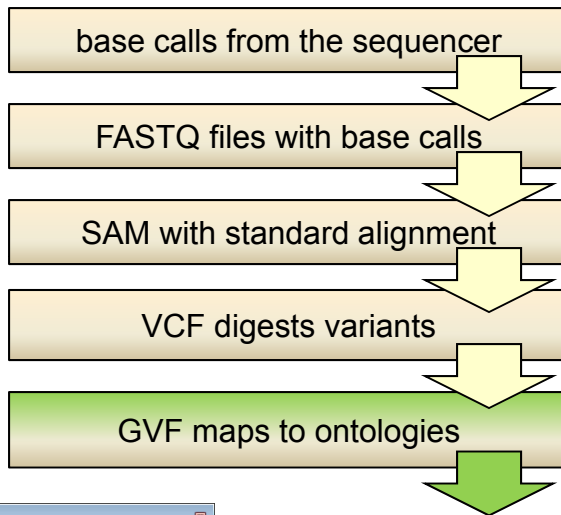
Inventing new Big Data Architecture



Big Picture - Data flow of next-gen sequencing

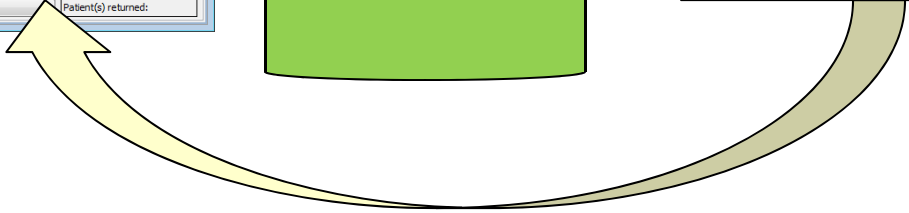
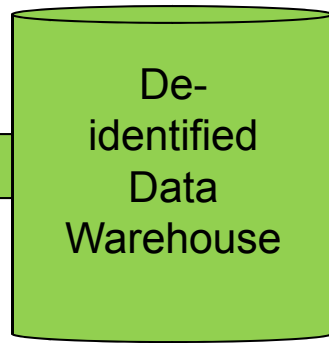
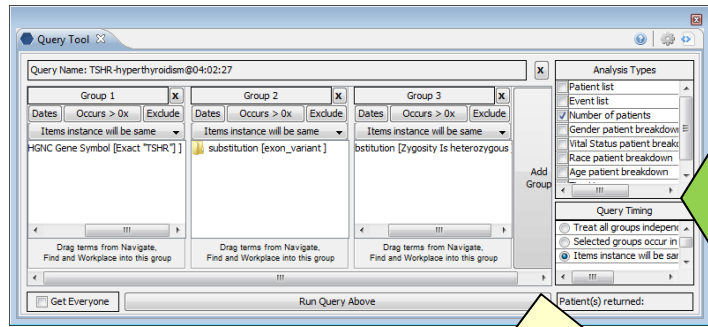


Data flow of next-gen sequencing

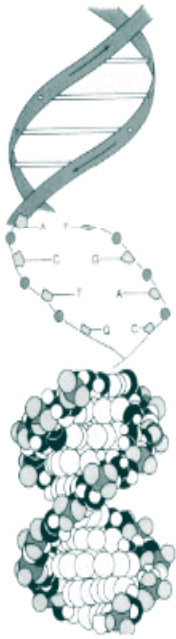


Map-Reduce Queries

Sequence patterns



I2b2, SHRINE, and SMART Information and Software on the Web



i2b2 Homepage (<https://www.i2b2.org>)

i2b2 Software (<https://www.i2b2.org/software>)

i2b2 Community Site (<https://community.i2b2.org>)

SHRINE at Harvard (<http://shrine.catalyst.harvard.edu>)

SHRINE Software:

(<https://open.med.harvard.edu/display/SHRINE/Software>)

SMART i2b2 Homepage (<http://smarti2b2.org>)

SMART Platforms Homepage (<http://smartplatforms.org>)

References

- Friedman C, Hripcsak G, Johnson SB, Cimino JJ, and Clayton PD. A generalized relational schema for an integrated clinical patient database. In Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care, IEEE Computer Society Press, Los Alamitos, CA, 1990, pp 335--339.
- Kimball, R. The Data Warehousing Toolkit. New York: John Wiley, 1997.
- Murphy, S. N., Morgan, M. M., Barnett, G. O., Chueh, H. C. Optimizing Healthcare Research Data Warehouse Design through Past COSTAR Query Analysis. Proc AMIA Symp. 1999;892-897
- Murphy, S.N., Gainer, V.S., Chueh, H. A Visual Interface Designed for Novice Users to find Research Patient Cohorts in a Large Biomedical Database. AMIA, Fall Symp. 2003: 489-493.
- Murphy, S.N., Weber, G., Mendis, M., Gainer, V.S., Churchill, S., Kohane, I.S. Serving the Enterprise and Beyond with Informatics for Integrating Biology and the Bedside (i2b2). Journal of the American Medical Informatics Association, 2010 March 1; 17(2): 124-130.
- Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, Kohane, I.S. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 2009;16(5):624-30

Co-investigators, developers, and collaborators

■ RPDR

- Eugene Braunwald
- John Glaser
- Diane Keogh
- Henry Chueh

■ I2b2, SHRINE and SMART

- Isaac Kohane
- Susanne Churchill
- Griffin Weber
- Michael Mendis
- Vivian Gainer
- Lori Phillips
- Rajesh Kuttan
- Wensong Pan
- Janice Donahue
- William Simons (SHRINE)
- Andy McMurry (SHRINE)
- Doug MacFadden (SHRINE)
- Ken Mandl (SMART)
- Josh Mandel (SMART)
- Nich Wattanasin (SMART)

■ Medical Imaging (mi2b2)

- Christopher Herrick
- David Wang
- Bill Wang

■ i2b2 Driving Biology Projects

- Roy Perlis
- Jordan Smoller
- Dan Iosifescu
- Vivian Gainer
- Victor Castro
- Andrew Cagan
- Robert Plenge
- Stan Shaw
- Qing Zeng
- Guergana Savova
- Caitlin Clements
- Wouter Hoogenboom,
- Tianxi Cai



THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

**See you on June 18-20, 2013 at the 3rd
i2b2 - SHRINE Conference in Boston!**

