
PhenoDB: Phenotypic Data Repository for Genetic Research in Complex Disorders

MGH Clinical Research Retreat
Sept. 11, 2003

*Dr. David Pauls, Dr. Pamela Sklar, Dr. Susan Santangelo,
Heather Nicholson, Jes Fagerness, Usmanou Nsangou*

Psychiatric and Neurodevelopmental Genetics Unit
Massachusetts General Hospital
<http://pngu.mgh.harvard.edu>

Cal Collins, Nitin Sawhney, Ph.D. and Tom Hickerson

Akaza Research, Cambridge, MA
<http://www.akazaresearch.com>



Sept 11, 2003 : 1 :: 20



Outline of Talk

- Phenotypic Data Repository for Clinical Research
 - » Motivation
 - » Challenges of Clinical Research
 - » Approach and Deployment

- PhenoDB: Overview of Current Release
 - » Creation of User-defined Phenotypic Parameters
 - » Submission & Validation of Data
 - » Managing Multiple Sites and Projects in a Centralized Repository
 - » Secure Distributed Access to Data

- HIPAA and Security for Clinical Research Settings

- Summary of Preliminary Evaluation

- Future Directions



Sept 11, 2003 : 2 :: 20



Motivation: Collaboration on Clinical Research for Rare Disorders

- Key Objective: To gain insights on complex phenotypes that characterize rare disorders
- Clinical studies on genetic causes of rare psychiatric disorders such as Schizophrenia, Autism and Tourette Syndrome
- Collaborative research among investigators at Mass General Hospital and various academic institutes worldwide
- Studies require collection of both phenotypic (exhibited traits) and genotypic (genetic makeup) data on afflicted individuals & families



Sept 11, 2003 : 3 :: 20



Challenges for Databases in Clinical Research Settings

- Ensuring a **robust, scaleable and extensible system** for distributed usage at research sites worldwide
- Providing **intuitive** query, data entry and administrative tools for use by investigators and clinicians
- Generalizable for **diverse clinical studies** in a centralized database
- Enabling **bulk data import** from legacy sources like Access & KP5
- Ensuring **double validation** of data submitted and auditing
- Access for **hierarchical roles, privileges & data sensitivity levels**
- Compliance with **HIPAA regulations** & guidelines
- Collecting longitudinal data, integrating genotypic data and enabling statistical analysis



Sept 11, 2003 : 4 :: 20



PhenoDB: Phenotypic Data Repository - Approach

- Pilot Project: Support international consortium of researchers investigating Tourette Syndrome – led by Dr. David Pauls, Harvard Medical School
- Centralized repository to securely manage data and provide integrated data analysis for cooperative research (distributed online access)
- Import bulk legacy clinical data, integrate with genotypic data and export to various analytic formats
- Dynamically generate various existing and custom psychiatric evaluation instruments and query tools from predefined metadata
- Manage hierarchical projects and roles, while providing intuitive workflows
- Deployment:
 - » Sun Fire V480 Server running Solaris 8
 - » Oracle 9iR2 database
 - » Java J2EE framework with Tomcat webserver



Sept 11, 2003 : 5 :: 20



PhenoDB: Overview of Current Release

- Creation of User-defined Phenotypic Parameters
- Submission & Validation of Data
- Managing Multiple Sites and Projects in a Centralized Repository
- Secure Distributed Access to Data



Sept 11, 2003 : 6 :: 20



PhenoDB Interface: Data Submission and Query

The screenshots illustrate the PhenoDB interface for data submission and query. The top-left screenshot shows the login page with fields for 'User ID' and 'Password'. The middle-left screenshot shows the 'Query Library: Create or Select Queries' page, which includes a table of queries:

ID	Name	Description
<input type="checkbox"/>	007 test of new report format	looks for external project id
<input type="checkbox"/>	208 test of s001 and s1002	test of query for skings
<input type="checkbox"/>	215 test 25 and test 75	test of some outliers
<input type="checkbox"/>	222 00001 and 004 75	we will see if two and want both have 3 in the 000
<input type="checkbox"/>	223 looking at 00001 or 000 > 0	Looking at some of these traits may lead to false

The top-right screenshot shows the 'PhenoDB: Enter Phenotypic Data' page, which displays a confirmation message: 'Your changes were inserted into the database.' Below this, there is a list of data points:

- Project: TS-5b Par
- Family: 041
- Instrument: TS-5b Par Self Report v1.5.0
- Individual: 000-003-02
- Status: Currently being collected

The bottom-right screenshot shows a detailed view of a questionnaire item, 'Starting the good mood, how long is the longest period you have gone without a minute to...', with a dropdown menu set to '0 - always gone w/o'.

PhenoDB Interface: Data Validation

The screenshot illustrates the PhenoDB interface for data validation. It shows a message about discrepancies in validating the data: 'There are some discrepancies in validating this data. Please look below and update them as necessary.'

The main section is titled 'PhenoDB: Validate Phenotypic Data' and displays a list of data points:

- Project: TS-5b Par
- Family: null
- Instrument: ESACG VT v.1.2
- Individual: 010-003-02
- Status: already complete

The main section displays a questionnaire item, 'Starting the good mood, how long is the longest period you have gone without a minute to...', with a dropdown menu set to '0 - No Information'.

Below the questionnaire item, there is a text box for validation comments: 'Please write any validation comments in the textbox below. Please note: These comments will also propagate across all sections, so please indicate questions and pages where appropriate.'

PhenoDB Interface: Managing Clinical Studies

PHENOTYPIC & NEURODEVELOPMENTAL GENETICS UNIT

Home | Subject | Profiles | Collection | Add/Manage Families | Add/Manage Individuals | View Data Entry Status | Enter Data

PhenoDB: Status of Submission

Your changes were inserted into the database.

Section	Questions Entered	Questions in Section	Data Entered	Validation
Motor Tic Checklist	0	136	Not Completed	Not Completed
Phobic Tic Checklist	0	50	Not Completed	Not Completed
Description of Multiple Tics	4	4	Completed	Not Completed
Current Severity of Tic Symptoms	0	0	Completed	Not Completed
Worst Ever Severity of Tic Symptoms	4	4	Completed	Not Completed
Age At Onset	15	15	Completed	Not Completed
Obsessive Compulsive Checklist	0	461	Not Completed	Not Completed
Worst Ever Severity of Obsessive Thoughts and Compulsive Behaviors	0	5	Not Completed	Not Completed
Age At Onset, Obsessive Thoughts	0	5	Not Completed	Not Completed
Age At Onset, Compulsive Actions	0	5	Not Completed	Not Completed
Help Sought Obsessive Thoughts and/or Compulsive actions	0	7	Not Completed	Not Completed
Attention Deficit Hyperactivity Disorder Checklist	0	89	Not Completed	Not Completed
Medications Taken Currently or in Past	0	132	Not Completed	Not Completed

This instrument is **Incomplete** for this individual.

[Click here](#) to return to the Collect Data menu.

PhenoDB: Status of Submission

HIPAA Regulations and Standards

- Health Insurance Portability and Accountability Act of 1996 (HIPAA)
 - » Goals: Privacy, Security, Standardization and Efficiency in Healthcare*
- Administrative Simplification Provisions
 - » Electronic transactions and code sets standards
 - » **Privacy requirements**
 - » **Security requirements**
 - » National identifier requirements
- Key Challenge
 - » Providing reasonable Security & Privacy while ensuring adequate performance and minimizing administrative overhead?

* <http://www.cms.hhs.gov/hipaa/>



HIPAA Privacy Rule for Clinical Research

- Privacy Rule limits release of “protected health information” (PHI) without patient’s knowledge and consent
- Applies only to “covered entities” which may include healthcare groups, organizations & businesses that handle PHI
- Researchers may not be covered entities, unless they are also health care providers who electronically transmit PHI during studies
- PHI considered individually identifiable information e.g. Name, Date of Birth
- Covered entities may disclose PHI without restriction, if it is de-identified
- Covered entities should seek individual’s written permission for use of PHI for research purposes



Sept 11, 2003 : 11 :: 20



HIPAA Rule on Security Standards

- Security standards to safeguard the "confidentiality, integrity and availability" of electronic information used in health care
- Final rule published on Feb. 20, 2003 and only applies to covered entities*

Key Provisions:

- General Rule Provisions
- Administrative Safeguards
- Physical Safeguards
- **Technical Safeguards**
 - » Access control policies - unique user ID, auto logoff, encryption of PHI
 - » Transmission security - integrity controls & encryption (not over open networks)
 - » Audit controls – hardware, software and procedural methods
 - » Person Authentication – ensure person seeking access to PHI is who they claim

* <http://www.hipaadvisory.com/regs/finalsecurity/summaryanalysis.htm>



Sept 11, 2003 : 12 :: 20



Data Privacy and Security: Implications of HIPAA

- **Authentication**
- **Integrity**
- **Authorization**
- **Availability**
- **Confidentiality**
- **Auditing**



Sept 11, 2003 : 13 :: 20



Data Privacy and Security: Utilizing Oracle9iR2 Features

- **Authentication** —————> *SSL & Single Sign On*
- **Integrity**
- **Authorization** —————> *Oracle Label Security*
- **Availability**
- **Confidentiality** —————> *Selective Encryption*
- **Auditing** —————> *Oracle Auditing*



Sept 11, 2003 : 14 :: 20



PhenoDB: Oracle Privacy and Security Approaches

- **Selective Data Encryption**
- **Oracle Label Security**
- **Oracle Auditing**



Sept 11, 2003 : 15 :: 20



Selective Data Encryption

- For Clinical Research, Encryption useful for:
 - » User Passwords
 - » Place of Interview
 - » Date of Birth (DOB)
 - HIPAA considers DOB individually identifiable, but researchers require it for their clinical studies
 - System computes age from DOB and Interview Date for access to most investigators
 - System provides DOB only to Principle Investigator, hidden from all other users including DBA
 - » Selective Encryption of Genotypic Data?



Sept 11, 2003 : 16 :: 20



PhenoDB: Using Oracle Label Security

Key Security Policies

- » Hierarchical Roles for Privileges to Clinical Studies & Instruments
- » Selective & Hierarchical Access to Data within different Project Sites

Label Components

- » *Levels*: Shareable, Internal, Confidential
- » *Compartments*: TS-Boston, TS-Utah, TS-London, TS-Toronto etc.
- » *Groups*: Project Director, Principle Investigator, Investigator, RA etc.

Combining label components to support complex security policies

- » Project Directors of TS-Boston can only set privileges for users in Boston, and access/update all levels of information.
- » Principle Investigators in TS-London can only set privileges for their own RAs, and only access/update Sharable and Internal information.
- » Investigators in TS-Utah can view sharable information from TS-Toronto



Sept 11, 2003 : 17 :: 20



PhenoDB: Auditing for Clinical Research Database

- Monitoring Access Problems & Potential Abuse
 - » Failed logins and access with non-existent users
 - » Attempts to access database at unusual hours
- Monitoring Changes to Clinical Protocols
 - » Trigger Audit Reports for instruments created/updated
- Monitoring Data Query, Submission & Validation
 - » Fine-grained audits for SQL queries matching prior predicates e.g. specific phenotypic parameters
 - » Triggers for data submission beyond normative range of responses
 - » Triggers based on data updates for validation of clinical data



Sept 11, 2003 : 18 :: 20



Summary of Preliminary Evaluation

- PhenoDB Pilot Evaluation Setup
 - » Evaluation Period: Aug 18th - Sept 10th
 - » Evaluations with 6 RAs entering data from 2 TS-OCD Self Reports
 - » Provided minimal training and familiarization

- Preliminary Results
 - » All completed the first phase of data submission
 - » Average time spent was 45 minutes in one or two sessions
 - Range: 20 mins to 8 hours
 - » Second case entered was easier than first
 - » Relative to previous approach, almost all users reported PhenoDB to be more user friendly, manageable, convenient and accessible

- Lessons Learned and Key Design Changes
 - » Training and familiarization is key to using the system
 - » When is an instrument considered complete?
 - » System should be flexible to not require filling all fields and use defaults
 - » Users moving to other sections before saving data – improve interface



Sept 11, 2003 : 19 :: 20



Future Directions

- Setup SCID, KSADS and other psychiatric instruments used in Tourette Syndrome Studies

- Implement other studies and projects at MGH

- Improve Advanced Query and Reporting Interface

- Allow bulk upload of data and complex metadata

- Implement Security Protocols and Policies

- Support longitudinal studies

- Integrate with Genotypic Data and Analytic tools



Sept 11, 2003 : 20 :: 20



Summary

- Distributed collaboration in clinical research settings present unique requirements & challenges for databases, particularly:
 - » Managing hierarchical roles, complex privileges & selective data access
- PhenoDB: Centralized repository to manage diverse clinical studies across multiple projects and sites, with integrated data access
- Compliance with HIPAA Regulations using Oracle Security
- Ongoing Evaluations and Iterative Interface Design
- Generalizing and scaling the system for diverse clinical studies
 - » Collaborating with MGH, Yale, U. Connecticut, UCLA and Rochester

