Response to the
Department of Health and Human Services

Request for Information on

Improving Health and Accelerating Personalized Health Care through Health Information Technology and Genomic Information in Population- and Community-based Health Care Delivery Systems

Submitted by

The Harvard Medical School-Partners HealthCare
Center for Genetics and Genomics (HPCGG)

On

January 10, 2007

Document Version

1.0

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Table of Contents

Abstract 1
Background – Challenges and Opportunities 1
The Promise of Genetic and Genomic Based Personalized Medicine 1
Improving Care with Genetic/Genomic Testing 1
Technological Advances Accelerating Adoption 2
The Challenges Facing Clinicians 3
Data that is Relevant throughout a Patient's Lifetime 3
Large Numbers of Discrete Data Points 4
The Effect of the Internet on Clinician-Patient Interactions 4
The Potential Opportunity for Clinical Researchers 5
The Importance of Biomarkers 5
The Role of Privacy-Protected Clinical Genomic Data 5
Enhanced Sample Availability-Privacy Protection 6
Genetics Related Information Technology Goals by Constituent 6
Partners HealthCare IT Solutions 9
The Partners IS/HPCGG Model Genomics Information Technology Architecture 9
Infrastructure that Can (and Should Be) Shared Across Clinical and Research Environments 10
Enterprise Laboratory Information Management System Superstructure 11
The Role of Laboratory Information Management Systems 11
Enterprise LIMS Superstructures 12
Our Gateway for Integrated Genomics-Proteomics Applications and Data 13
Phenotype-Genomic Variant Correlation Databases 14
The Utility of Phenotype-Genomic Variant Correlation Databases 14
GeneInsight 15
The Potential Value of Real-Time Updates 15
Challenges with Providing Direct Access to Physicians 16
Consent Tracking System (CTS) 17
Infrastructure Support for Genetics and Genomics in Clinical Medicine 18
Genomic Variant Interpretation Engine (GVIE) 21
The Electronic Medical Record 22
Privacy Enhancements 23
User Interface Enhancements 23
Optimizing Storage of Genetic/Genomic Data 23
Support for Genetics Based Clinical Decision Support 27
Inter-organizational Communication and the Role of Standards 29
Infrastructure Support for Genetics and Genomics in Clinical Research 31
Distributed Bioinformatic Analysis Environment (I2B2 Hive) 31
Research Patient Data Registry (RPDR) 32
Virtual Specimen Locator (VLS)/Shared Pathology Information Network (SPIN) 33
Prospective Sample Collection Engine (Crimson) 34

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

RelNet 35
The Importance of Protecting Patient Privacy 35

Government Intervention in Specific Areas Could Accelerate the Development of Personalized Medicine IT 36

Safeguarding Privacy 36
Resolving Privacy verses Clinical Care Tradeoffs 37
Clarifying Legal Responsibilities (and Limits Thereof) 38
Providing a Basis for Interacting with Personal Health Records 39

A Unified-Idealized Scenario 40

Patient Visit 40
The Diagnostic Testing Laboratory 41
The Research Process 41
The Laboratory 41
The Clinical Decision Support System (CDSS) 42

Conclusion 42

References 43

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Abstract

The incorporation of genetic knowledge and genomic technologies into clinical medicine ("personalized medicine") has the potential to change clinical outcomes and radically improve medical practice. The revolutionary effect these technologies will have on the health care system is already being felt: genetic tests are being used to guide treatment in clinical domains as diverse as heart disease, cancer, infectious disease, and many other common illnesses. But we are still early in this revolution. As the movement toward genetic and genomic informed personalized medicine accelerates, substantial new pressures will be placed on our health care delivery system. Many of these pressures are addressable through sophisticated enterprise and inter-enterprise IT solutions. The Partners HealthCare Information Systems Department, the Harvard Medical School-Partners HealthCare Center for Genetics and Genomics (HPCGG), and Hewlett Packard Corporation have been collaborating for more than three years to enhance Partners HealthCare's IT infrastructure to meet the needs of personalized medicine. The collaboration focuses on enhancing enterprise information technology infrastructure that supports research and clinical environments. This document will describe the infrastructure we have deployed thus far as well as additional enhancements we believe are necessary to prepare us for wider adoption of genetic and genomic informed personalized medicine. We will also suggest ways the federal government may be able to assist in the development of this type of critically needed infrastructure.

Background – Challenges and Opportunities

The Promise of Genetic and Genomic Based Personalized Medicine

Improving Care with Genetic/Genomic Testing

Improved patient health is the goal of personalized medicine. Personalized medicine seeks to empower clinicians by providing tools that predict susceptibility to common diseases, determine the prognosis for those diagnosed with a particular disease, and use genetic/genomic information to determine the treatment that is most likely to benefit the patient. With this

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

information, physicians and other clinicians can make better informed choices among treatment options. Because virtually all diseases, except trauma, have a genetic component [1], genetic and genomic based diagnostics are a key part of the vision of personalized medicine. Many medical conditions commonly thought of as single diseases are actually complex: many different diseases that happen to present similar symptoms.[2] Hypertension [3] and asthma [4] are but two examples of such complex disorders. To treat the root cause of a patient's disease and take into account factors that could modify the efficacy of different treatments, it is important to determine the role of the patient's genetic composition in the onset and progression of the disease, as well as the response to therapy. Molecular diagnostics, also known as genetic or genomic based diagnostics, provide clinicians with the means to determine disease-specific genetic and genomic factors in any patient. Once these features are identified, clinicians can consider correlations that have been established between those specific features and treatment efficacies. For example, we offer molecular diagnostics to:

- determine whether non-small cell lung cancers have genetic variants that correlate with either Tyrosine Kinase Inhibitor (TKI) efficacy or resistance
- determine whether patients have genetic variants that trigger cardiac anomalies that lead to sudden death if left untreated
- identify whether a patient's hearing loss is caused by genetic variants that are also correlated with other serious medical conditions
- identify whether a patient has genetic variations that will cause him or her to metabolize Warfarin abnormally, either quickly (risking stroke) or slowly (risking brain hemorrhage).

Technological Advances Accelerating Adoption

The field of clinical molecular diagnostic testing is evolving quickly. A few years ago, nearly all genetic tests were gene-based tests that involved examining a small number of specific base pairs in a patient's DNA to determine whether particular mutations or variants were present. Today, we commonly run sequencing tests that read long segments of patient DNA and determine all variations present in those sequences. Chip-based technologies are now breaking through into the clinical arena, which will make it possible to survey increasingly large segments of DNA in a cost-effective manner. There is also increasing optimism that expression profile

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

arrays, which simultaneously measure the level of expression of all known genes in the human genome, will revolutionize personalized medicine. The NIH is funding research aimed at bringing down the cost of sequencing a person's entire genome to approximately $1,000. While we are a few years away from reaching this goal, research will continue to drive down the cost of sequencing. Newer methods based on detecting proteins or metabolites in body fluids are also expected to revolutionize the ability to predict disease onset, progression, or treatment.

While the cost of sequencing each base is going down, the size of the clinical sequencing market is rapidly increasing. In 2002, the genetics portion of the molecular diagnostics market was only $16 million. By 2005, it had grown to $94 million and is projected to reach $423 million by 2009.[5] When considered in combination with the decreasing cost of sequencing, this implies that the amount of patient DNA being sequenced is growing exponentially and with it the number of genetic variants being identified in the patient population. While this information will prove very helpful to clinicians, it will also present them with significant new challenges.

The Challenges Facing Clinicians

Data that is Relevant throughout a Patient's Lifetime

There are several important differences between genetic and genomic test results and most other forms of clinical data. It is estimated that one individual differs from another by as many as a million genetic variants. Many millions of such variants have already been described and the number will increase. Even with incomplete knowledge, no clinician will be expected to know what role each of these genetic variants plays in health or disease. This problem becomes more complex as we learn new information about the variants. Although genetic variations found through molecular diagnostic tests on germ line DNA (DNA that is representative of what was inherited from the patient's mother and father) will not change over a patient's lifetime, our understanding of the clinical implications of the variants will increase. This represents a challenge for clinicians. Currently, when a molecular diagnostic test result is delivered, it comes with an interpretation in the form of a text report. Therefore, the clinician does not have to do specific research into the implications of the variants identified in the report to ascertain their meaning. The challenge comes as these interpretations may change. How will clinicians gain access to new discoveries on variants found in their patients? Clinicians' time is already fully allocated; there is no

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

slack in the system that would allow them to continuously research the latest discoveries associated with every variant found in every patient.

Large Numbers of Discrete Data Points

Genetic results are largely discrete. Sequencing test results produce lists of variants. This differs from laboratory results that take the form of a quantitative value that lies along a continuum. Clinicians can gain a sense of the implications of a finite number of quantitative values over time. However, it is more difficult to get a sense of genetic data. Multiple variants in a gene may have unique clinical implications, each with its own probability of contributing to a disease. It would be impossible for a person to comprehensively track the implications of even a small portion of the variants that could be found in his or her patient population. Expression profiling data will present an even greater challenge. Expression profiling chips can generate more than 45,000 quantitative data points. A human cannot possibility manually extract meaning from these data; algorithms are required. How will updates to these algorithms be applied to pre-existing patient data?

The Effect of the Internet on Clinician-Patient Interactions

The challenges associated with tracking the implications of genetic variants will increase as the number of variants identified in our patients grows exponentially. The challenge will be further intensified as patients themselves increasingly use the internet to conduct their own genetic research. In 2004, the last year for which data is available, the median amount of time physicians were able to spend with their patients in a typical visit was 14.7 minutes.[6] What will happen when patients begin to bring printouts to their appointments, either touting the virtues of new genetic tests or describing the new and potentially worrisome implications of the patient's genetic variants? How will clinicians evaluate the quality of the information, explain the appropriate course of action to the patient, and accomplish everything else that needs to be taken care of in the median visit time of 14.7 minutes? Genetic and genomic data, in combination with the internet, has the potential to greatly increase the time pressure on clinicians. If this pressure is not appropriately managed, clinical care could degrade. We must give clinicians the tools they will need to fully leverage the benefits of genetic and genomic data while protecting their time.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

The Potential Opportunity for Clinical Researchers

The Importance of Biomarkers

Much of the clinical research community is focused on genetic and genomic biomarker discovery. Biomarkers are observable/assayable physical properties that correlate with disease proclivities, disease subtypes, drug metabolic rates, or other clinical characteristics relevant to the practice of medicine. They form the basis of molecular diagnostic tests and, therefore, the basis of genetic and genomic based personalized medicine. Biomarker discovery also identifies new drug targets for the pharmaceutical industry.

The rate of biomarker discovery is highly influenced by the costs associated with obtaining the relevant genetic, genomic, and clinical data. Currently, to identify a human genetic or genomic biomarker, a clinical researcher must identify a set of patients that either has been or can be clinically phenotyped, and then obtain samples from those patients that can be subjected to genetic analysis. This process of obtaining a clinical cohort and related samples is usually time-consuming and expensive. After obtaining the samples, the researcher will almost always need to incur additional costs associated with the genetic and genomic analysis.

The Role of Privacy-Protected Clinical Genomic Data

The movement of genetics and genomics into clinical medicine has the potential to dramatically reduce the costs of cohort selection and targeted sample acquisition. As described above, an increasing amount of genetic data is being generated in the clinical setting. Genomic data is likely to follow. Providing this data to clinical researchers can increase the efficiency of biomarker discovery in multiple ways. There is the possibility of retrospectively analyzing genetic, genomic, and other phenotypic data gathered through the clinical process to search for new correlations. These new correlations could identify a new biomarker or enhance the understanding of an existing biomarker. Because the data are gathered through clinical processes, researchers can perform this type of analysis without incurring the costs associated with obtaining and analyzing samples. The amount of genetic data currently gathered through clinical care is still small, so it is difficult to perform this type of analysis today. With increasing adoption of genetic testing, however, it is likely that this technique will become viable for a rapidly increasing number of analyses.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Even the focused genetic clinical data currently collected has the potential to help clinical researchers. These genetic data points can be combined with clinical data to stratify populations and produce focused research cohorts. These patients can then be contacted regarding their willingness to participate in research studies.

Enhanced Sample Availability-Privacy Protection

Enhanced biosample banking infrastructure could also provide significant benefits to the clinical research community. Biosample banks could facilitate access to samples that can be subjected to additional genetic analysis to substantially augment the data collected through clinical processes.

Of course, protecting patient privacy is of critical importance to all the research processes discussed in this section. (Please see "The Importance of Protecting Privacy" section for a more detailed discussion of this issue.) Internal Review Boards (IRBs) will not allow this type of research to proceed unless proper safeguards are in place. Physicians and patients themselves will not feel comfortable storing genetic data in patient medical records if they do not feel their information will be adequately protected. Information systems can play a critical role in establishing mechanisms that enable clinical researchers to access clinically derived genetic information in a secure, privacy-aware, and audited manner. They can also establish mechanisms that facilitate the sharing of samples across institutions while enabling each institution to maintain the control it needs to protect privacy. (Please see "Virtual Specimen Locator (VLS)/Shared Pathology Information Network (SPIN)" section for more information.)

Genetics Related Information Technology Goals by Constituent

Substantial IT support will be required to realize the promise of personalized medicine, address the challenges genetic data will present to clinicians, and capture the benefits these new technologies could bring to clinical research. There are multiple constituencies that will require expanded IT support.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Table 1: Genetic/Genomic IT Requirements of Each Constituency Group

Constituency	Requirements
Clinicians	• Quick access to: o summaries of the clinical utility associated with specific molecular diagnostic tests (both to determine which tests a particular patient could benefit from and to educate themselves about tests patients have seen on the internet) o molecular diagnostic test reports o information on the clinical significance of variants that have been identified in each of their patients • Clinical decision support to: o provide warnings when an action being taken (for example, prescribing a drug) is contraindicated by a patient's genetic profile o assist in determining optimal medication and its dosage o alert clinicians when new information becomes available on a variant that could materially improve patient care o alert clinicians when a molecular diagnostic test becomes available that may materially benefit specific patients • Active notification of test results (as some tests have turn-around times in excess of a week)

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Constituency	Requirements

Clinical Researchers	• IRB regulated, privacy-protected access to clinically derived genetic, genomic, and phenotypic information
	• Search functions capable of identifying what patient samples are available for further analysis
	• Unified study management system that allows them to:
	  o login and create a study
	  o manage who is allowed to access information associated with a study
	  o add samples to the study
	  o order additional clinical, genetic, genomic, and proteomic analyses on these samples through one user interface that allows them interact with all relevant laboratories
	• Access to databases with variant-specific interpretations
	• Access to computational and storage resources suited for working with genetic data
	• Integration of clinical, laboratory test, genetic test, and image data

Laboratory Personnel	• IT support for laboratory processes focused on work coordination, quality control, and process automation
	• Integration with enterprise systems to facilitate:
	  o clinical and research result reporting
	  o inter-laboratory workflow handoffs
	  o interactions with customers
	  o financial processing
	• Integration with databases of known variants and other genetic data that can be used to optimize laboratory processing

Geneticists	• Access to and the ability to curate genetic and genomic databases that can assist in results reporting • Systems that automate the result-reporting process
Patients	• Access to their complete medical record in a format that could be taken to another provider if necessary • Security and privacy safeguards that prevent their information from being used inappropriately • Systems that coordinate their care across physician interactions and ensure their genetic and genomic profiles properly inform clinical decision making

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Partners HealthCare IT Solutions

The Partners IS/HPCGG Model Genomics Information Technology Architecture

The Partners HealthCare Information Systems Department (Partners IS), the Harvard Medical School-Partners HealthCare Center for Genetics and Genomics (HPCGG), and Hewlett Packard (HP) have been collaborating for more than three years to develop an integrated IT architecture capable of supporting genetic and genomic based personalized medicine. This architecture is designed to support the needs of the constituent groups outlined in Table 1. We have found that enterprise IT infrastructure and applications are required to meet these needs. Therefore, we have extensively leveraged Partners' existing enterprise information architecture in both the research and clinical domains. Much of the architecture is highly developed and routinely used by clinicians and researchers; several additional components are under development; and a few areas are still in the planning stages. However, the entire architecture is expected to evolve quickly as the use of research and clinical genetic and genomic techniques accelerate. We offer this architecture as a model for others to evaluate, critique, and improve upon. Figure 1 shows the high-level research and clinical application stacks that form the basis of our architecture.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

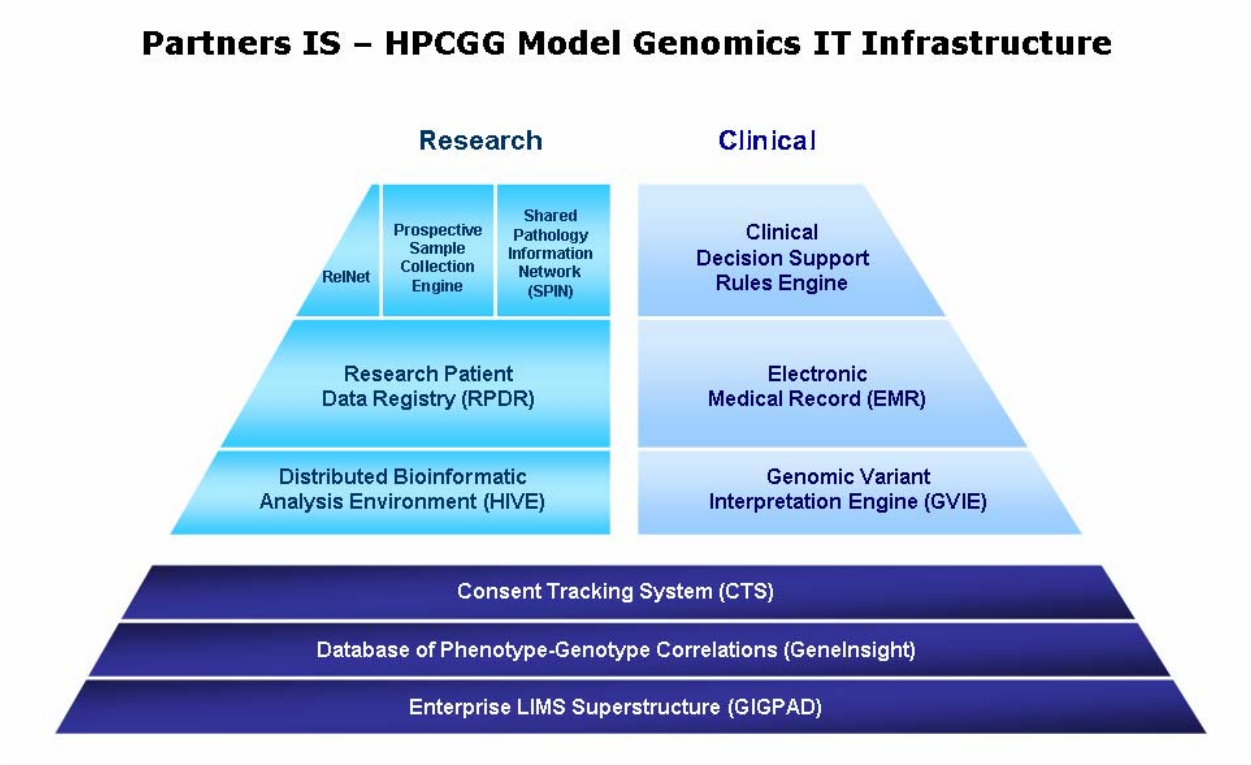

Figure 1: Partners IS/HPCGG Model Genomics IT Infrastructure.

Infrastructure that Can (and Should Be) Shared Across Clinical and Research Environments

At the enterprise level, three common functions are shared between research and clinical organizations that focus on leveraging genetics and genomics. Both must assay samples to obtain genetic and genomic data. Both are interested in developing knowledge bases that record the significance of genetic and/or genomic variants. Both would also benefit if the consent management process were handled in a structured, unified manner. We have found it useful to recognize these similarities between research and clinical organizations and build enterprise systems designed to support both environments. The concept of "translational medicine" leading to personalized medicine depends on being able to efficiently migrate processes and procedures from the research environment to the clinical environment. Enterprise IT systems that span across these two environments can simplify and speed the process of adding support for new clinical technologies and techniques when they breakthrough from the research environment.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Enterprise Laboratory Information Management System Superstructure

The Role of Laboratory Information Management Systems

Almost all genetic and genomic data are generated in laboratories. Laboratory Information Management Systems (LIMS) are critical to an overall genomics IT infrastructure because they can capture genetic and genomic data in structured form when it is initially generated. Downstream bioinformatics, report generation, and clinical decision support systems depend on this structured genetic and genomic data. LIMS are also important for ensuring data integrity across the inter-organizational process flows associated with genetic testing. For these reasons, integrated LIMS support is an essential part of a genomic IT enterprise architecture. In addition, LIMS can help reduce costs and increase quality through process automation, facilitating communication and reducing the need for manual entry of information.

A large number of genetic and genomic technologies are used in research; while the number used clinically is smaller, it is growing. Different technologies tend to require different types of LIMS. Therefore, an institution that operates a variety of genetic and genomic laboratories will need to support multiple LIMS, each of which will have an independent life cycle: as new technologies are introduced, new LIMS may need to be installed; as old technologies are retired, it is at times possible to decommission the LIMS that supported them.

This creates a challenge for enterprise integration efforts. System interfaces are expensive to build and maintain. Maintaining a large number of such interfaces to individual laboratory LIMS would be cost-prohibitive for most institutions. Furthermore, individual LIMS often do not have the robust security and user management infrastructure required of systems that are integrated into the enterprise. End users on the research side think in terms of “studies” that may leverage data generated in multiple facilities. Research users should not have to learn and manage multiple user interfaces and passwords to interact with the different facilities they use to generate data. Creating an Enterprise LIMS Superstructure can help address these challenges. (It is also possible that such LIMS systems exist at different locations and no enterprise LIMS superstructure could be developed. The solution in these cases is to develop standard data formats that can be exchanged across heterogeneous LIMS environments.)

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Enterprise LIMS Superstructures

An Enterprise LIMS Superstructure serves as an umbrella over the individual LIMS in the environment. Individual LIMS are integrated with the superstructure and then the superstructure is integrated with other systems in the enterprise environment. The superstructure also exposes a single, consistent user interface that external users leverage to interact with any of the underlying laboratories. For this reason, the superstructure handles all authentication and authorization of users outside of the laboratory. The individual laboratory LIMS only needs to authenticate laboratory users and therefore, can be isolated at the network level from the wider institution. We have found that the superstructure concept works fairly well because, while internal processes vary considerably across laboratories, the way laboratories interact with outside users is surprisingly consistent. The superstructure focuses on supporting the types of processes that are relatively consistent while delegating the variable processes to the specialized LIMS. Because the processes that involve interactions with the outside environment tend to be less variable, the interfacing challenge is simplified.

An Enterprise LIMS Superstructure has three other key advantages. First, because a considerable amount of code is moved out of individual LIMS into a common component, individual LIMS become smaller and easier to maintain. This reduces costs and also makes it faster to establish LIMS support for new laboratories. Second, we have found, thus far, that the differences between research and clinical environments can be largely isolated in the superstructure. This means that the individual LIMS becomes relatively agnostic as to whether it is working in a research or clinical environment. Clinical and research instances of the superstructure can be created that share a common code base. This should enable LIMS to quickly transition from the research to the clinical environment when a clinically significant breakthrough occurs. IT can contribute to translational medicine by facilitating this type of reuse. Third, the superstructure can support workflows that involve multiple laboratories and facilitate handoffs and data integrity across these processes. We have found this capability particularly important in the clinical environment.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Our Gateway for Integrated Genomics-Proteomics Applications and Data

In our environment, we created a system called the Gateway for Integrated Genomics-Proteomics Applications and Data (GIGPAD) to serve as our Enterprise LIMS Superstructure. GIGPAD also serves as a platform for constructing custom LIMS for laboratories when commercial and open-source offerings are insufficient. GIGPAD has been in operation since April 2004 and currently supports seven facilities. We are actively developing support for an additional three facilities. Utilization has grown considerably after the system was introduced. On November 6, 2006, there were 542 GIGPAD registered users. The growth in assays tracked by GIGPAD is shown in Figure 2.

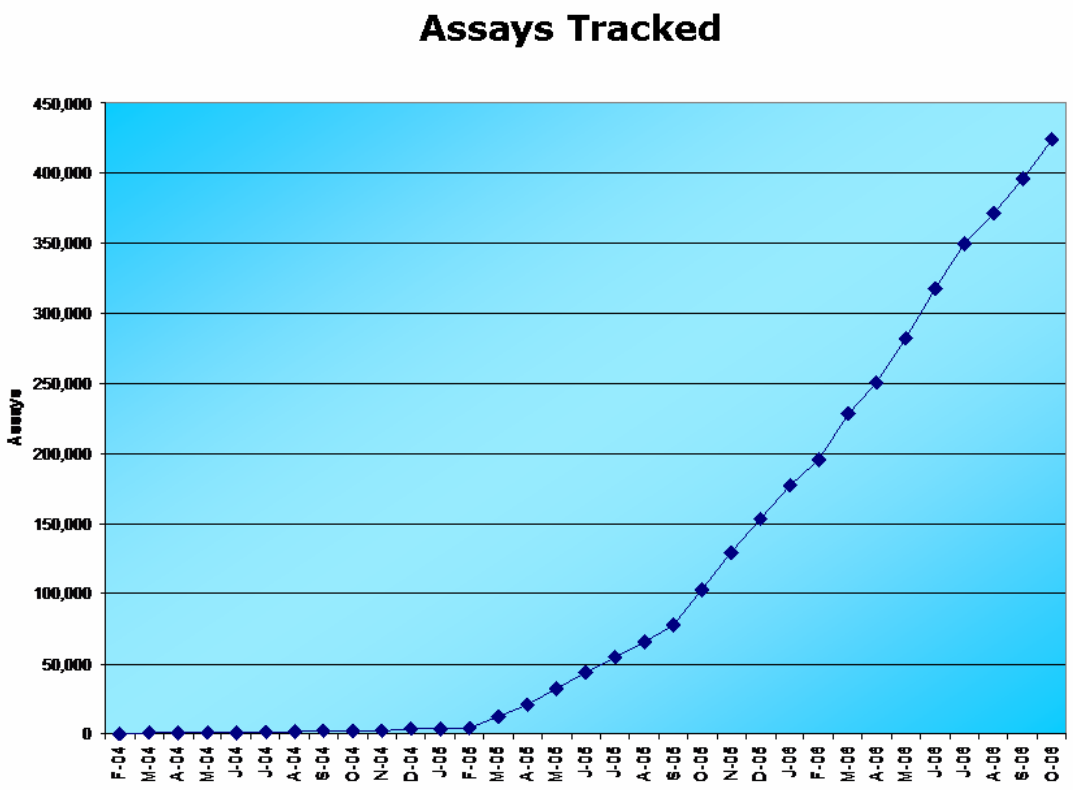

Figure 2: Growth in assays tracked by GIGPAD over time.
Note each sequencing well is counted as an assay.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

As the research technologies supported by the HPCGG break through into clinical use, we will have an opportunity to test the portability of our LIMS across environments.

Enterprise LIMS Superstructure
Key Functions

- Serves as a single interface point for all encapsulated LIMS
- Provides a common user interface leveraged by external users to interact with laboratories
- Handles authentication and authorization of external users
- Coordinates workflows that involve multiple laboratories
- Supports both research and clinical environments
- Manages processes that do not vary significantly across laboratories

Phenotype-Genomic Variant Correlation Databases

The Utility of Phenotype-Genomic Variant Correlation Databases

The practice of genetic and genomic-based personalized medicine is built on a constantly expanding knowledge base of phenotype-genomic variant correlations. When a correlation is discovered between a clinically significant fact and a genomic variant, a molecular diagnostic test can be developed. The test then provides clinicians with a means of determining whether a variant is present in their particular patient. Our lung cancer, heart disease, hearing loss, and pharmacogenomic molecular diagnostic tests all arose from the discovery of phenotype-genomic variant correlations.

Because genomic knowledge is constantly expanding, our understanding of the clinical implications of particular variants can change over time. Therefore, it is important to have a database that tracks current knowledge relative to individual variants. As with other aspects of our genomic IT infrastructure, we anticipate that the need for this type of database will grow as the amount of patient DNA sequenced, and therefore the number of genomic variants identified in patients, also increases.

Keeping this type of database current is extremely challenging. Geneticists working in molecular diagnostic laboratories tend to be extremely busy and are usually unable to dedicate significant time to maintaining these

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

databases for the diseases they study. Within the research community, there are a large number of heterogeneous databases where correlations are recorded. These databases range from large centralized repositories to small websites maintained by individual laboratories. Many of these databases are not intended or structured for clinical use. The variation across these databases will make it extremely difficult to keep a centralized resource continuously updated.

GeneInsight

We have developed a database, called GeneInsight, that currently addresses some but not all of these challenges. GeneInsight has data structures that associate information with diseases, genes, tests, and genetic variations. When a new test is brought on line, developers load data from existing data sources. GeneInsight is then integrated into the geneticists' reporting processes so that it is maintained as a by-product of the process of signing out reports. This is made possible through integration with the Genomic Variant Interpretation Engine (GVIE) described in the next section. As a result of this process, geneticists can maintain GeneInsight for the diseases they report on without a significant incremental time investment when they encounter a new variant. As a benefit, the time required to report on previously identified variants is significantly reduced. Overall, the combined GVIE/GeneInsight system saves geneticists time, which promotes systems utilization.

The Potential Value of Real-Time Updates

Updates to GeneInsight made through laboratory curation do not always occur in real time relative to the knowledge discovery. Usually, they are triggered through the identification of patients with particular variants in the laboratory. Because GeneInsight is updated when a variant is identified in a patient, it is sufficient to meet the results reporting needs of a laboratory. However, GeneInsight is also a knowledge repository that we intend to ultimately integrate with our clinical decision support systems. Real-time updates to GeneInsight would be preferable for clinical decision support purposes because, ideally, new knowledge would be leveraged to improve patient care as soon as it becomes available. Because the discoveries driving these "knowledge events" could occur anywhere in the world and be posted to any one of a number of databases, we would like to integrate GeneInsight with as many sources of variant-specific genetic information as possible. To begin the process we are working on developing a generic

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

integration process for interfacing instances of the Leiden Open Variation Database (LOVD) with GeneInsight. LOVD is an open-source database that researchers can use to publish information related to specific variants.

Ideally, it would also be possible for clinical laboratories to distribute the information contained in their Phenotype-Genomic Variant Correlation databases. This information is curated for clinical use and therefore represents a useful addition to information exposed by the research community. The information curated in a clinical laboratory would be of use to other clinical laboratories as well as health care providers. We are interested in exploring models for federating GeneInsight instances installed across institutions as an additional way to meet this need. Doing so would entail distributing GeneInsight to health care providers as well as other clinical laboratories. If clinical laboratories integrate GeneInsight into their reporting processes they would build knowledge bases as they reported on variants. These knowledge bases could then be federated either by establishing a central hub designed to interconnect versions of GeneInsight or enabling instances of GeneInsight to communicate with each other in a peer-to-peer fashion.

Challenges with Providing Direct Access to Physicians

Ultimately, we would like to provide clinicians direct access to the interpretations contained within GeneInsight. A number of issues must be resolved, however, before this can be done. These include assigning responsibility for keeping the information up to date and ensuring that clinicians do not lose track of the clinical context in which the variant resides. Geneticists, pathologists, and other health care professionals apply professional judgment when they generate reports. Mechanisms must be developed that enable variant-level information automatically to be delivered to clinicians while at the same time enabling clinicians to identify situations where the expertise of a geneticist and/or pathologist is required to develop a properly updated interpretation.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Database of Phenotype-Genomic Variant Correlations Key Functions

- Provides a structured means of tracking the relationships between:
  - Genetic Variants
  - Genes
  - Diseases
  - Clinical Interpretations
  - Molecular Diagnostic Tests
  - References
- Integrates with reporting systems to maximize utility and facilitate maintenance
- Serves as a knowledge base for clinical decision support systems
- Integrates with other sources of genetic or genomic information

Consent Tracking System (CTS)

A patient can enter and travel through an academic medical center in a number of ways. For this reason, it is difficult to track patient consents with any detail at the enterprise level. At Partners, the consent process begins with a uniform HIPAA notice signed by all patients seen by a provider, plus additional consents collected for specific studies and/or procedures. In our organization, the process of contacting patients to inquire about their interest in participating in studies is generally mediated through the patients' physicians. It is often difficult for busy physicians to perform this role. This contributes to the great time and expense associated with recruiting patients for research studies. From the patient's perspective, their exposure to clinical trials and other research depends on their physician's willingness to facilitate these interactions. The resulting variability with different physicians may be suboptimal for the patient. Also, the lack of enterprise architecture leads to the development of multiple consent tracking systems operating at the study level. This increases both study expenses and the variability of consent tracking procedures across studies.

For all these reasons, we believe an enterprise Consent Tracking System (CTS) would have significant benefits. With the proper training of clinical staff, patients could be asked for their consent in the clinical environment

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

and given the option of indicating whether they would be interested in participating in specific types of studies. They could also be given the option of specifying whether they would like to contribute samples for research purposes and, if so, what those samples can be used for. Clinical researchers would then have the option of requesting IRB approval to contact patients or obtain samples based on the information in the CTS system. The process for handling situations where a patient withdraws consent could be handled centrally through unified IT infrastructure.

Establishing a CTS would be an expensive undertaking. Substantial IT and non-IT costs are involved. We are beginning the process of evaluating the feasibility of such a system and potential funding options.

Consent Tracking System Key Functions

- Enable patients to specify whether they would be willing to be contacted about specific types of studies
- Enable patients to specify whether they would be willing to contribute samples to specific types of studies
- Record consents obtained from patients in a central database
- Provide patients a single point of contact they can leverage to withdraw any or all consents given to the institution

Infrastructure Support for Genetics and Genomics in Clinical Medicine

From a clinical provider perspective, there are multiple personalized medicine process flows that will require support. On June 20, 2006, we established support for the first of these process flows (shown in Figure 3). This flow results in structured genetic test results flowing from the HPCGG Laboratory of Molecular Medicine to a specially secured portion of the Partners HealthCare Electronic Medical Record when a Partners patient is tested. Based on this support, we are now beginning the process of initiating support for the flows shown in Figure 4. For completeness, Figure 5 shows interactions with external sources that will be needed for optimal

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

functioning of the IT infrastructure. The remainder of this section discusses the dedicated clinical components of our information architecture.

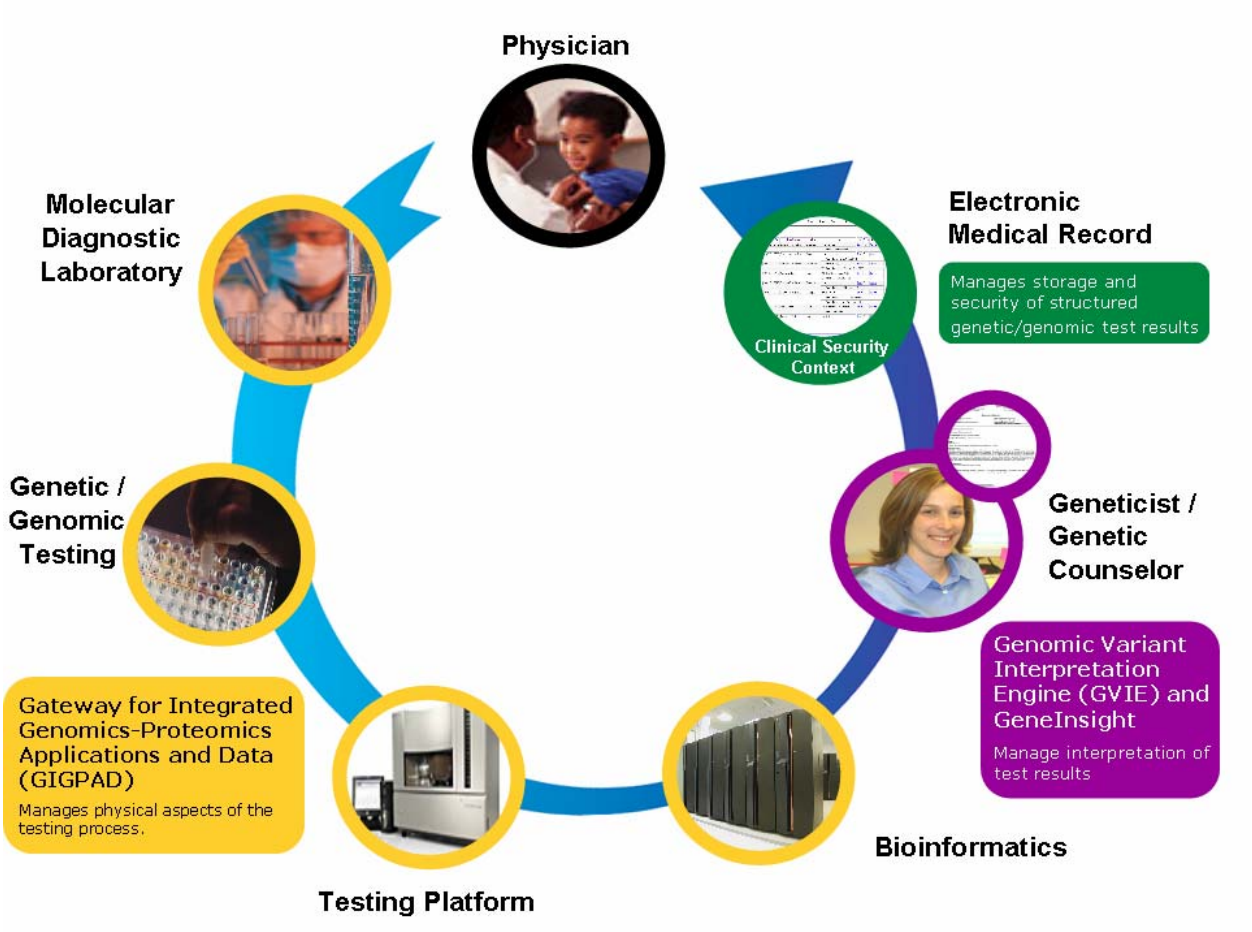

Figure 3: Information technology that currently links the HPCGG Laboratory for Molecular Medicine with Partners HealthCare Electronic Medical Record. This infrastructure results in structured genetic data integration into the Electronic Medical Record.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

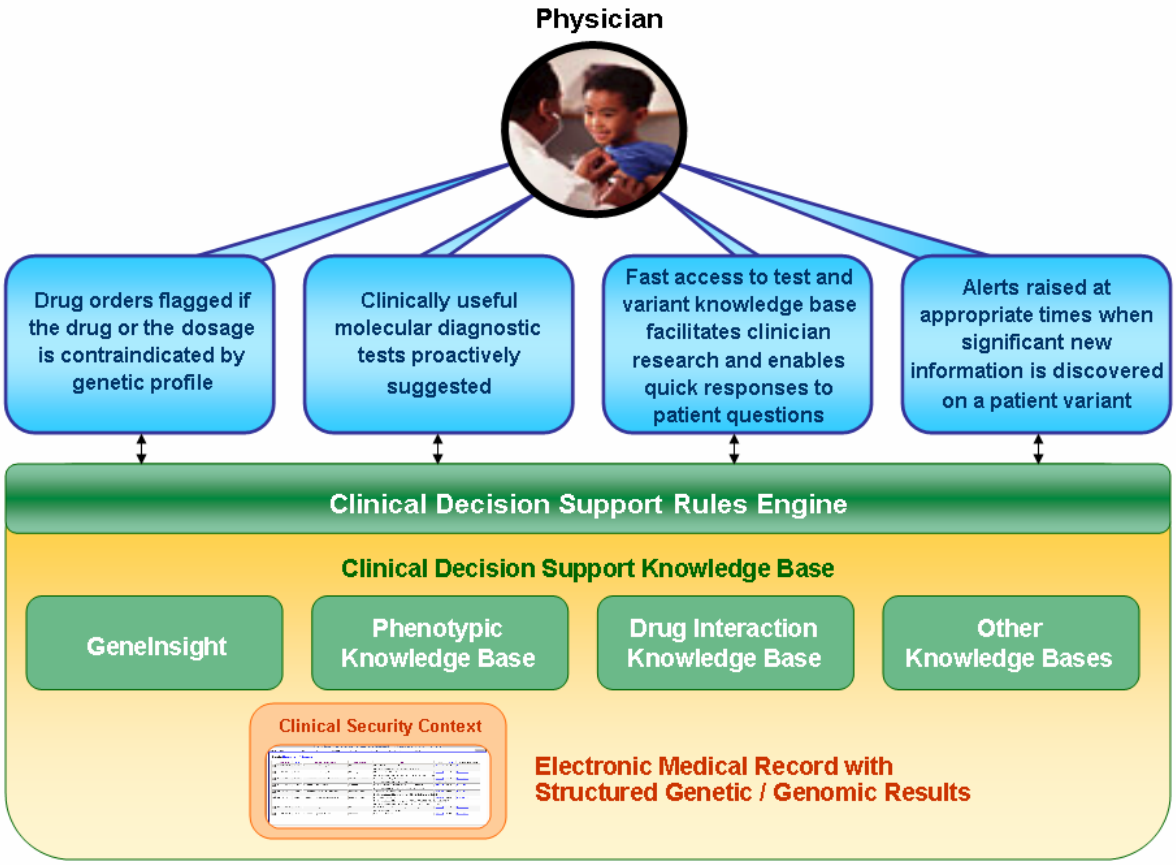

Figure 4: Example of clinical decision support functions (the top 4 boxes) that could leverage genetic data in the EMR.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

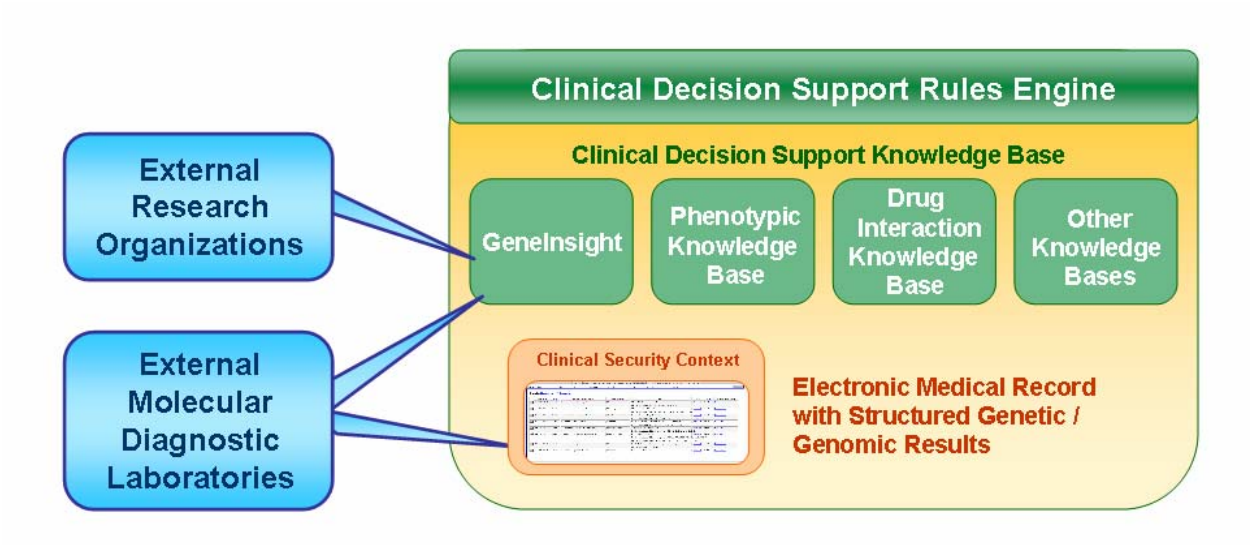

Figure 5: External process flows required to fully support genetic/genomic based personalized medicine. External molecular diagnostic laboratories must supply genetic/genomic test results in structured form to EMR infrastructure for the data to be leveraged by clinical decision support rules engines. Both external laboratories and research organizations could supply knowledge associated with genetic/genomic variants.

Genomic Variant Interpretation Engine (GVIE)

Most clinicians have neither the training nor the time to assess the clinical significance of variants that have been identified in their patients. For this reason, molecular diagnostic laboratories typically employ geneticists who interpret physical test results and produce a text report describing the significance of any genetic variants identified. The process of generating this report can be time-consuming and expensive, so streamlining and automating portions of the process through IT can be valuable. IT can also help standardize result reporting by reducing variability between the ways different geneticists might interpret the same result. When test results are sent to the EMR, it is useful to capture interpretations in structured form in addition to the genetic variants themselves. Capturing structured interpretations requires IT support during report generation.

To meet these needs, we have constructed a genetic results reporting tool we call the Genomic Variant Interpretation Engine (GVIE). Our geneticists enter the variants found in a patient into GVIE and then GVIE queries GeneInsight to determine if the significance of the identified variants has already been determined. If so, then GVIE formulates a draft report based

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

on the information stored in GeneInsight. If one or more of the variants is not known to GVIE, the geneticist is guided through the process of entering the relevant information for that variant into GeneInsight and a draft report is generated based on this new information. Either way, a geneticist reviews the draft report, makes modifications or additions if necessary, and signs the case out. Both the unstructured text report and the structured variant and interpretation information then flow to the EMR.

As GVIE matures, it may be able to support alternative process flows that transfer responsibility for result interpretation from geneticists to clinical domain experts. For example, we may encounter situations where experts in one of our clinical domains express a desire to specify and control the clinical significance of a set of variants relative to their domain. We could expose GVIE and GeneInsight to these domain experts and allow them to take control of the interpretation process. Some have suggested that clinical associations will ultimately take responsibility for variant interpretation. If this comes to pass, we will work out mechanisms for importing rules from those associations into GVIE.

Like GeneInsight, today GVIE is only used during the report-generation process. Yet there maybe situations where a clinician might want to use GVIE to see how a set of variants identified in a patient a long time ago would be interpreted today. We would have to develop procedures for validating GVIE for use by non-geneticists before this could happen, but we would ultimately like to provide this type of functionality

Genomic Variant Interpretation Engine (GVIE)
Key Functions

- Automates process of genetic / genomic report generation
- Integrates with database of phenotype-genomic variant correlations to obtain data for interpretations
- Integrates with electronic medical record to deliver results

.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

The Electronic Medical Record

Privacy Enhancements

The scope of electronic medical records (EMRs) extends far beyond genetics and genomics. However, there are a few important enhancements that are required to make an EMR handle genetic and genomic information. The most important of these relates to security augmentation. Partners HealthCare maintains a Clinical Data Repository (CDR) as part of its EMR infrastructure. The CDR stores test results. We found it important to "section off" a portion of the CDR to create a genetic/genomic profile area that could be specially secured. While all data in the CDR needs to be protected and secure, there are certain forms of genetic tests that require heightened security. Genetic tests can reveal a predisposition for disease before actual symptoms emerge. These types of results are considered particularly sensitive. Massachusetts has enacted specific regulations to heighten protections surrounding these types of screening results. Our CDR genetic profile module only allows ordering physicians to see genetic results derived from molecular diagnostics tests that were classified as "screening tests." In addition to building this logic into our CDR, we also need to ensure that all other applications that access the data are subject to the same privacy protections. This was done by exposing the data through "service" interfaces. These services require an authenticated username, which they use to filter results and exclude data the user should not see. (A genetics non-discrimination law could also help protect the privacy of data in the EMR.)

User Interface Enhancements

In addition to enhancing privacy protections, we also needed to add pages to our CDR to display genetic results. Genetic results are different than other results in that they are often valid for a patient's lifetime, so it is important to create user interfaces that would enable clinicians to scan through large numbers of accumulated test results. Figure 6 displays one such interface. We decided to build the interface so that the individual rows can be expanded to show which specific genetic variants were identified through a test. Only very specialized clinicians with in-depth knowledge will be able to leverage this variant information. However, we felt it was important to provide for the needs of this type of clinician. At the same time, exposing these variants makes it clear that this information is being stored in the CDR and is available for use in clinical decision support algorithms. These clinical

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

decision support algorithms could, in turn, provide wider-scale benefits. A clinician who is not familiar with the meanings of individual variants can click on a test to be presented with the interpretative report provided by the geneticist. Figure 7 displays a sample of such a report.

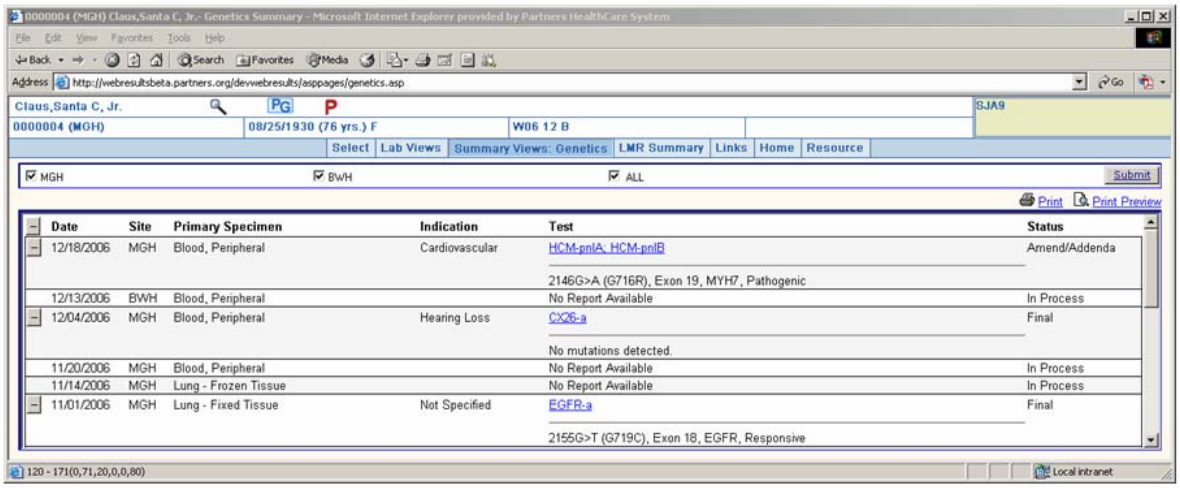

Figure 6: Partners HealthCare Clinical Data Repository (CDR) genetic test review screen.

Optimizing Storage of Genetic/Genomic Data

When we integrated our Molecular Diagnostics Laboratory with our CDR, we had to decide what information would be stored in the CDR and what information would reside in other systems. The sequencing-based genetic testing process involves the creation of a sizable quantity of "raw" instrument files that are then processed through automatic and manually executed bioinformatic processes. The ultimate output of these bioinformatic processes is a list of genetic variants that can be represented in fully structured form using short text strings. Therefore, the "raw" instrument files are much larger than the processed output. There is an argument that raw instrument files should be stored in the EMR because there is always a chance that variants were incorrectly extracted and that improved future bioinformatic algorithms may be able to leverage old raw instrument files to catch these mistakes. However, we chose not to store these raw instrument files in the EMR and instead required the enterprise LIMS superstructure (GIGPAD in our case) to store and track these files over time. This eases the burden on the EMR while providing for long-term storage of the underlying data. A common accession number serves as a unique identifier that preserves the linkage between tests in the CDR and the laboratory records in GIGPAD. The full list of variants derived from these raw instrument files is sent to the CDR in fully structured form.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

HARVARD MEDICAL SCHOOL | PARTNERS HEALTHCARE

Laboratory for Molecular Medicine
Center for Genetics and Genomics
65 Landsdowne Street, Cambridge MA 02139
Phone: (617) 768-8500 – Fax: (617) 768-8513

Unit Number(s):
Lab Accession: **PM-06-F00768**
Patient Name: **CLAUS, SANTA B.**
Birth Date: **5/15/0006**
Age Sex: **2000 Year old Male**

MOLECULAR DIAGNOSTICS REPORT

Specimen Type:	Lung - Fixed Tissue	**Report Date:**	12/07/2006
Related Accession(s):		**Received Date:**	12/7/2006
Referring Physician:	XXXXXXXXXXXX	**Referring Facility:**	MGH
Copies To:		**Referring Fac. MRN:**	
		Lab Control Number:	

GENETICIST - XXXXXXXXXX

TEST PERFORMED - EGFR-a

TEST DESCRIPTION - EGFR Partial Gene Sequencing (Exons 18-21)

INDICATION FOR TEST - Adenocarcinoma

RESULTS

DNA VARIANTS:
2369C>T (T790M), Exon 20, EGFR, Resistant
2573T>G (L858R), Exon 21, EGFR, Responsive

INTERPRETATION:
Resistant. The T790M mutation in combination with the XxxxX mutation has previously been described in an individual with acquired resistance to EGFR-TKIs (Pao, et al., 2005).

TEST INFORMATION

BACKGROUND:
Somatic mutations in the kinase domain of the epidermal growth factor receptor (EGFR) have been shown to predict response of tumors to tyrosine kinase inhibitor (gefitinib and erlotinib) therapy in approximately 80% of patients with advanced lung non-small cell carcinoma.

METHODOLOGY:
DNA was isolated from the dissected tissue and the nucleotide sequence encoding the four exons of the kinase domain (exons 18-21) of EGFR was analyzed by PCR and capillary gel electrophoresis. When necessary, testing is also performed on the blood specimen to determine if an identified variant is specific to the tumor or present constitutively. This test was developed and its performance characteristics determined by the Harvard Medical School/Partners Healthcare Laboratory for Molecular Medicine (LMM, 65 Landsdowne Street, Cambridge, MA 02139; CLIA#22D1005307). It has not been cleared or approved by the U.S Food and Drug Administration (FDA). The FDA has determined that such clearance or approval is not necessary.

REFERENCES:
Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med. 2005 Feb 24;352(8):786-92.

xxxxxxxx

xxxxxxxxxxxx
www.hpcgg.org
CLIA# : 22D1005307

Figure 7: Example report

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

As additional genes are implicated in diseases, molecular diagnostics tests for these diseases may be expanded to cover these additional genes. This increases the sensitivity of the molecular diagnostic tests. For this reason, it is useful to record the exact regions of DNA that were sequenced in any given molecular diagnostic test. This information could be leveraged by future clinical decision support systems to help determine cases where it makes sense to order additional genetic tests on a patient. We store test definitions in GeneInsight, which fully describes the regions of DNA sequenced. A unique test ID is sent to the EMR to maintain a link to this test definition.

Variants identified in the patient are represented differently to the clinical users and computer algorithms. Our EMR displays human readable forms found in literature. GVIE translates results according to testing platform, and formats these human readable display lines into "display line" tags in the results message. The CDR leverages the content of these tags in its displays. This design simplifies the process of adding support for new testing platforms that may require new human readable result formats. For example, encapsulating results in this way enabled us to add support for genotyping results in October 2006 without an EMR code change. (The June 2006 release of EMR included support for sequencing, but not genotyping, results.)

Interpretations can be associated with individual variants or with a test as a whole. Both types of structured interpretations, along with the text report, are stored in both the CDR and GVIE. Table 2 displays the systems that store various data elements generated through the molecular diagnostic testing process. Our infrastructure currently only supports sequencing and genotyping based genetic test results. We believe our strategies will remain largely consistent when expression profiling techniques come into clinical use, although we will need to make a decision whether to store all or part of the gene expression values in the CDR. As an alternative, we could store these values in GIGPAD along with the raw instrument files and just store the interpretations in the CDR. Our decision will depend on how we assess the likelihood that expression datasets will require reinterpretation over time.

Table 2: Storage locations of different types of data generated during the genetic testing process.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

File	GIGPAD	GVIE	GeneInsight	CDR
Raw Instrument Files	√			Reference: Accession Number
List of Genetic Variants	√	√		√
Test Definition			√	Reference: Test ID
Structured Interpretations (including display line tags)		√		√
Text Report		√		√

Electronic Medical Record Genomic Extensions Key Functions

- Enhanced security to protect sensitive genetic/genomic tests
- Interfaces to laboratories that collect and store structured variant information for use in clinical decision support
- User interface screens optimized to display genetic/genomic information

Support for Genetics Based Clinical Decision Support

Genetic and genomic based clinical decision support is the end goal of much of our genomics IT infrastructure. As displayed in Figure 4, this decision support will take many forms. Some forms of decision support may involve direct propagation of updates from GeneInsight (genetic "knowledge events") to clinicians. Other types of clinical decision support will involve combining genetic or genomic data with phenotypic data. For example, we are conducting a clinical trial focused on using genetic data to develop a better Warfarin dosing algorithm. If we are successful, we will make this algorithm available to clinicians through a Warfarin dosing calculator. The inputs to this calculator will include phenotypic (such as height, weight, age) and genotypic information (such as P450 and vitamin K receptor variants).

We believe there are at least two critical components to a sustainable and scaleable genetics-aware clinical decision support system. First, a

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

genetic/genomic knowledge repository (in our case GeneInsight) must be installed and accessible alongside knowledge bases covering other areas. Second, a clinical decision support rules engine must be present. Because genomic knowledge is dynamic, optimally leveraging this knowledge will likely involve frequent additions or changes to the clinical decision support rules base. It is important that these changes be easy and inexpensive to make. We believe a clinical decision support rules engine could make it possible to add or update clinical decision support rules through a graphical user interface without involving application programmers or needing to integrate it into software release processes. The reduction in costs would be dramatic.

We are currently exploring whether a third piece of infrastructure is required to handle "knowledge events" – situations where new knowledge becomes available on variants previously identified in patients. There is an open question as to who should be responsible for identifying and propagating these knowledge events. Laboratories may not have the bandwidth to do so, especially given that they may report on a geographically dispersed patient population managed by a large number of providers. If the laboratory does not manage these updates, providers will need to generate them internally or gather them from an external source. Either way, they may want to update existing reports containing outdated interpretations. They will need a system similar to GVIE to do this, though one that can operate without geneticist intervention. This system would need the ability to apply updates in mass to large numbers of text reports and their accompanying structured interpretations. If such an infrastructure existed, then rules engines could rely on the interpretations present in the EMR without needing to contain logic to constantly re-derive them.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Genomics Aware Clinical Decision Support Key Functions

- Rules engine to simplify rule construction and maintenance
- Interfaces to genetic/genomic as well as phenotypic knowledge repositories
- An automated report update engine may also become important

Inter-organizational Communication and the Role of Standards

The GIGPAD to GVIE/GeneInsight to CDR integration described above enables the transfer of structured genetic tests results to our EMR when a Partners patient is tested in our own Laboratory for Molecular Medicine. However, like all health care providers, Partners uses external reference laboratories for a great deal of our genetic testing. Similarly, our Laboratory of Molecular Medicine performs many genetic tests for external providers. For structured test results to transfer in these cases, interfaces will need to be established between laboratories and provider organizations as depicted in Figure 5. In addition to the need to receive structured information from other institutions, ideally genetic knowledge and CDSS rules would be as portable as possible. Common data structures, coding standards, and ontologies would greatly assist these efforts.

We have been working with several standards bodies (i.e., HL7 Clinical Genomics Special Interest Group, LOINC, and NCBI) coming to understand existing baseline models and contributing modification needs (from lessons learned during Partners implementations). Additionally, we collaborate with leaders in these organizations for knowledge transfer of implemented messages, vocabularies, test dictionaries, and corresponding use cases on an ongoing basis, as well as suggesting guideline updates. (These individuals represent the following institutions: Healthcare and Life Sciences Standards Practice, IBM Haifa Research Lab; Intermountain Healthcare, Salt Lake City, UT; and Regenstrief Institute, Indianapolis, IN.)
We hope these efforts will ultimately lower the costs of establishing interfaces between institutions and ensure that data transferred is fully leveraged to improve clinical outcomes.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

We are currently collaborating with Intermountain HealthCare to establish an inter-institutional interface for passing structured genetic results. This effort includes collaboration on standards work with HL7 v2 modeling and LOINC. The interface will enable us to further test these standards, as well as potentially serve as a prototype for other organizations.

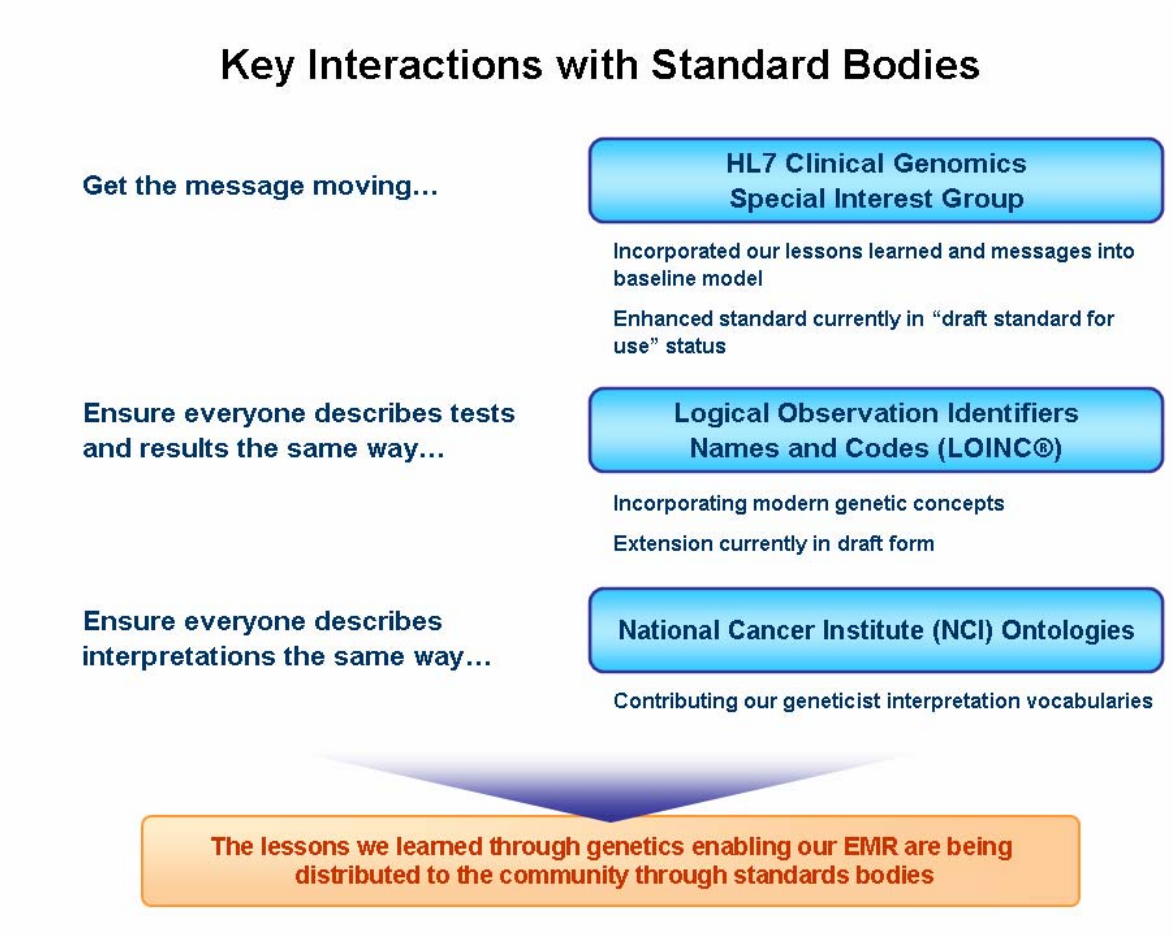

Figure 8: Interactions with Standards Bodies

The above interface efforts are focused on facilitating the transfer of genetic test results. As described above in the "Database of Phenotype-Genomic Variant Correlations" section, there is also a need to facilitate the movement of genetic and genomic knowledge between institutions. We hope to contribute to efforts to meet this need by developing a federated version of

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

GeneInsight that is capable of interfacing to heterogeneous knowledge sources as well as other instances of itself.

Infrastructure Support for Genetics and Genomics in Clinical Research

Enterprise IT infrastructure is becoming increasing important to clinical research organizations. As described above, enterprise LIMS architectures can contribute to laboratory efficiency. In addition to laboratory-focused systems, enterprise IT can help researchers access data and samples generated in the clinical environment. Constructing this type of infrastructure requires tight integration between research and clinical systems. Opportunities are also emerging to integrate bioinformatic routines into enterprise IT systems and give them access to centrally managed hardware resources.

Distributed Bioinformatic Analysis Environment (I2B2 Hive)

Typical bioinformatic analysis efforts involve bioinformaticians using scripting languages to link together a series of open-source modules to perform a particular function for a particular study. With the exception of very high-throughput laboratory processes, it is rare for these analysis pipelines to be routinely reused in different contexts throughout an enterprise. The Integrated Informatics Bench to Bedside (I2B2) National Center for Biomedical Computing located at Brigham and Women's Hospital in Boston (a founding member of Partners HealthCare) is creating a bioinformatic environment designed to address this issue. This environment, know as the I2B2 Hive, focuses on defining a mechanism for creating bioinformatic "cells." Every cell is capable of performing a bioinformatic task and also exposing interfaces to its functionality, which can be leveraged by other cells. Because of the modular nature of these cells, they should be able to be rearranged and combined to meet a variety of different enterprise needs. These cells will be combined with the Research Patient Data Registry (RPDR) described in the next section to create a fully functional clinical research chart.

Additional information regarding the I2B2 hive is available at: https://www.i2b2.org/resrcs/hive.html

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Research Patient Data Registry (RPDR)

Clinical data is of enormous value to clinical research efforts. However, this data is normally difficult and expensive for clinical researchers to obtain because of the need to secure and protect patient privacy. Only an IRB can give researchers the right to view identified patient data. This creates an awkward situation for researchers, one in which they typically need to go through an IRB approval process just to find out if there are enough patients in the system to form a large enough cohort for a study. Even if a researcher gains IRB approval, it can be difficult to extract data from clinical systems in a structured form suitable for querying. Clinical systems are usually transitionally focused, making it difficult to query them for research purposes.

We addressed these issues by constructing an IRB-approved Research Patient Data Registry (RPDR). The RPDR is constructed on a relational database (SQL Server) and interfaced to our clinical systems to obtain data, usually on a nightly basis. This is a large and ongoing effort and significant resources are dedicated to constantly expanding these interfaces. For example, diagnostic genetic data (but not screening data) can now be loaded into the RPDR. A user interface is exposed, which allows Partners researchers to query the database to determine roughly how many patients are in the system that meet certain criteria (we cannot provide exact counts because doing so would enable a sophisticated user to identify a patient.[7]). Once a researcher has identified a cohort, he or she can request IRB approval to receive identified data on that cohort and contact the patients' physicians to discuss enrolling them in clinical trials.

The RPDR has proven to be a very useful tool for our researchers. Since its launch in 2002, the scope of the RPDR continues to grow, and now includes not only patient demographic data, diagnoses and procedure data, pharmacy data, inpatient and outpatient encounter information, provider information, and laboratory data – but also data from the electronic health record (LMR). The RPDR has about 1,300 users throughout the Partners Healthcare System. The RPDR database is composed of more than 3.5 million patients and 900 million coded records from patient encounters, labs and results, and other medical care. Each coded event is represented as a fact in the database and in turn is associated with other important contextual information. Since its inception, a total of 2,155 identified data sets containing 10,069,224 patients have been returned to RPDR users. We estimate that in 2005, sponsors funded grants of $20.7-$30.7 million that

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

where critically dependant on the RPDR. The total funding of these critically dependent grants receiving data in 2005 ranged from $94-$136 million. I2B2 is now targeting the data in the RPDR through its correlation identification engine, described later in this section. We are now expanding the RPDR's functionality to give it the ability to query for samples as well as patients. This functionality will integrate with the other tissue repository efforts described below.

Additional information about the RPDR can be found at: www.mgh.harvard.edu/rcc/rpdr.htm

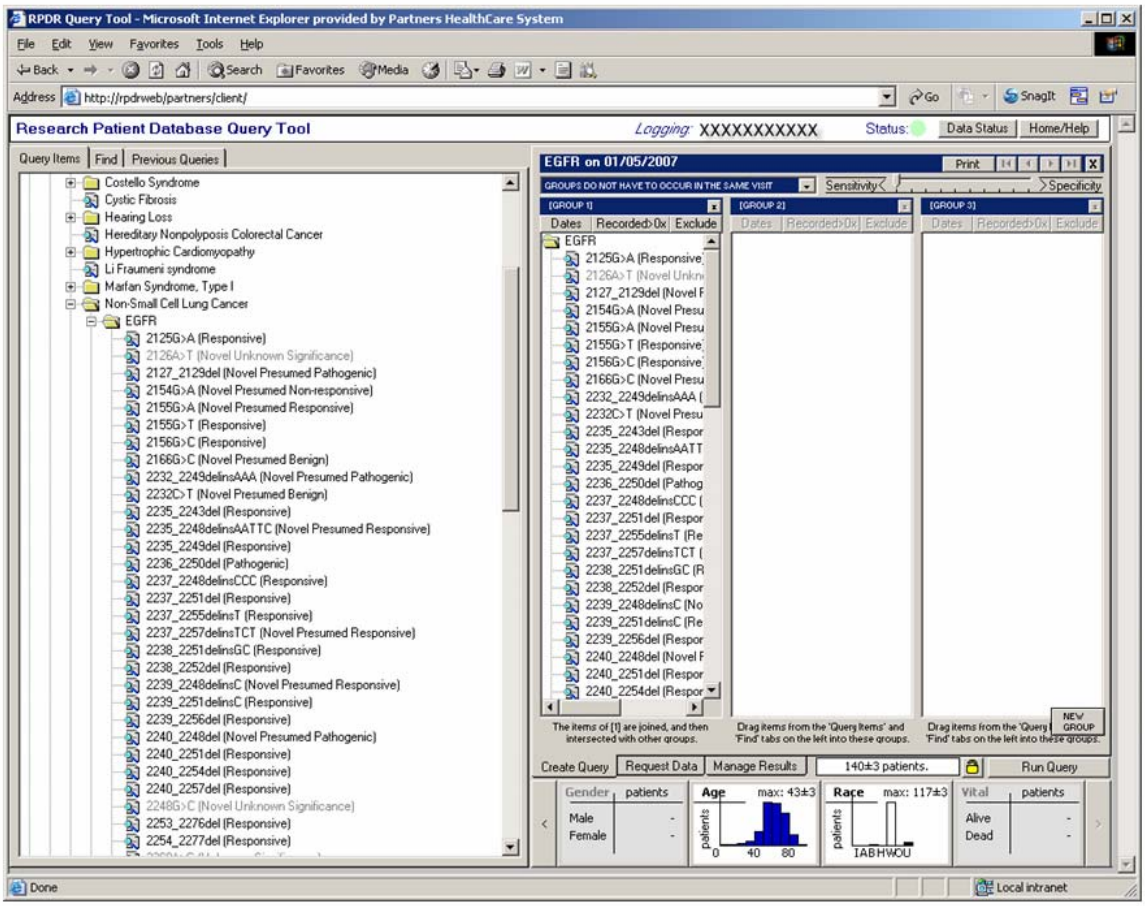

Figure 9: An example screen from the RPDR.

Virtual Specimen Locator (VLS)/Shared Pathology Information Network (SPIN)

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Pathology departments process a great deal of samples. At times, excess tissue or blood remains after a clinical process is completed. In these circumstances, pathology departments are allowed to "bank" these samples. With appropriate IRB approval, researchers can obtain these samples from the bank for research purposes. However, most clinical researchers have no way to find samples that might benefit them. The Shared Pathology Information Network (SPIN) is intended to offer researchers a resource that would allow them to search multiple banks simultaneously to locate samples that would materially benefit their research. SPIN aims to create a national infrastructure of data-sharing while maintaining local control. That is, there is no requirement for a central database but instead each participant manages his or her own autonomous database that participates as a node in the SPIN network. At any time, any institution can choose to withdraw from the network. However, precisely because this option is available, institutions are much more willing to join and participate so that millions of samples are made available on the SPIN network. Once these samples are identified, the clinical researcher can contact the directors of the banks where the samples in question are stored to discuss the possibility of using them in their research. SPIN is intended to operate across institutions. Each institution links itself to the peer-to-peer SPIN network by exposing a Virtual Specimen Locator (VSL). The VSL links together individual banks within an institution. For researchers within the HIPAA entity, the VSL can operate using a richer set of search criteria. Only de-identified information can be exposed to SPIN. We are currently in the process of architecting a VSL to serve the Partners infrastructure, and plan to integrate with the RPDR to leverage its query interfaces. Once this interface is complete, within our HIPAA entity, genetic information stored in our RPDR could be used in the query criteria for samples.

Additional information about the VSL is located here:
https://querytool.med.harvard.edu/querybuilder/search?action=about

Additional information about SPIN is located here:
http://www.cancerdiagnosis.nci.nih.gov/spin

Prospective Sample Collection Engine (Crimson)

In effect, the VSL/SPIN network gives users the ability to query existing sample inventories. This is a form of retrospective sample collection – it

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

targets samples that have already passed through the clinical process and have been selected for banking. It is also possible to prospectively select samples. We are currently working on a prospective sample collection system called Crimson. Crimson will allow researchers to establish criteria that define the types of samples that would benefit their studies. Crimson will interface with pathology systems and identify samples flowing through the clinical process that meet the criteria. If these samples are not entirely consumed during clinical processing, technicians are notified that they should be retained for the study instead of being discarded or banked. Crimson has facilities for associating anonymized clinical information with the sample. Using these facilities, researchers can obtain anonymized clinical samples with the clinical context preserved. In the case of IRB study-approved consented patients, identifying information can be retained. These samples can then be forwarded to systems like GIGPAD for genetic analysis.

RelNet

In an ideal world, we would be able to securely and consistently analyze the clinical data generated on our patients as it is produced to search for new and clinically useful facts. This would be the ultimate use of academic medical centers as living laboratories. New discoveries would become a by-product of the practice of medicine, even before additional resources are spent obtaining and analyzing samples. While the full implementation of this vision remains far in the future, we are taking steps toward this goal. Part of the I2B2 effort involves establishing a cluster that will build large relevance networks by correlating “everything against everything” in a database. Hewlett Packard has designed a cluster and storage solution to support this effort. The Hewlett Packard hardware together with the I2B2 software form the basis of the RelNet Project. RelNet will be connected to an instance of the RPDR that will be refreshed over time. As described above, the RPDR contains a very large amount of clinical data on a very large number of patients and therefore will enable the construction of sizable knowledge networks. Because the amount of genetic data in the RPDR is still small, it is unlikely that this type of analysis will initially identify phenotype-genotype correlations. However, as the amount of genetic data in the RPDR grows, these types of discoveries will become more likely.

The Importance of Protecting Patient Privacy

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

Protecting patient privacy is of paramount importance in all these efforts. No amount of genetic knowledge will help a patient who does not consent to be tested. There is a great deal of fear surrounding genetic testing. Patients are often concerned that genetic test results will affect their insurance, employment, or personal relationships. The issue of genetic privacy is further complicated by the fact that immediate family members have a 50 percent chance of having variants found in a patient.

Privacy concerns must be addressed at several levels. Federal and state governments provide input through regulations. Institutions can establish governance committees focused on this issue. At Partners, for example, we have established a genetics subcommittee under our privacy committee to steward these issues. IRBs provide guidance on privacy issues related to research use of data or samples. IT departments themselves also build functionality into their systems to protect privacy. For example, at Partners, we have incorporated functionality into our EMR to limit access to screening genetic test results to the ordering physician. All information stored in an EMR is confidential. Of course, all genetic results are subject to the EMRs authentication, authorization, and auditing mechanism. However, the presence of genetic and genomic information raises the bar and increases the importance of being vigilant in maintaining security and patient protections.

Government Intervention in Specific Areas Could Accelerate the Development of Personalized Medicine IT

Government legislation and/or regulation could speed the development of IT support for personalized medicine in several key ways. Government assistance would be particularly helpful in strengthening patient privacy protections while at the same time clarifying how trade-offs between patient privacy and clinical care should be resolved. It would be extremely helpful if the government could also help clarify legal responsibilities. Passage of a genetic non-discrimination law would provide protection to individuals about abuse of genetic/genomic information in employment.

Safeguarding Privacy

The government could help a great deal by strengthening regulations that safeguard patient privacy. Genetic data presents many important issues for

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

the insurance industry. When Dr. David Brailer was the National Coordinator for Health Information Technology he spoke about how the insurance industry relies on certain events being "unknowable" to both insurance companies and insured individuals. As genetics makes certain events more "knowable," the insurance industry will have to adjust. While this will become a major challenge, it is important that this challenge be separated from the health care delivery process. Health care providers must always be empowered to act in the best interests of their patients and be given legal standing to do so. In the case of genetic and genomic testing, health care providers should not be required to disclose either these results or the fact that a patient was tested – even if the patient signs a release – unless there are clear, federally established exceptions that are a response to a significant public interest. It is important that the House of Representatives take up debate on the Genetic Nondiscrimination Bill passed by the Senate (S.306). However, the insurance provisions in this bill focus on restricting the use of genetic information by health insurers. This is not sufficient to fully protect health care providers concerned about their patients' privacy. Additional protections need to be added to limit the situations where a health care provider can be compelled to disclose genetic information to outside parties – most notably life insurers. Issues regarding what tests the insurance industry can conduct should be resolved without the involvement of health care providers.

Resolving Privacy verses Clinical Care Tradeoffs

A number of privacy versus clinical care trade-offs will emerge as the clinical use of genetics and genomics expands. For example, at Partners, genetic information collected through genetic and genomic screening tests can be seen only by the ordering physician. This protects the patient's privacy but, over the long term, this rule has the potential to harm their care. A worst-case scenario would involve a screening test result that contraindicates a drug that might be used by an emergency room physician on an unconscious patient. This type of scenario has not arisen to date, but it could materialize at any time and it will be important that our infrastructure be constructed to react properly when it does. Should the emergency room physician be given special permission to see these types of test results? If so, are their other situations in which genetic privacy constraints should be relaxed to optimize care delivery? Who should make these decisions? In our case, these decisions are made by the genetics subcommittee of our privacy committee. As other institutions also make these types of decisions through their own

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

governance bodies, fragmentation and inconsistency may result. This will be harmful to patients because they will not know what to expect from each of the providers they deal with. Government could provide leadership in this area through principles, policies, and regulations to help reduce fragmentation and inconsistency.

Clarifying Legal Responsibilities (and Limits Thereof)

As described above, the structured nature and persistence of genetic data creates an opportunity for much deeper clinical decision support than is normally possible. This is fortunate, because for the reasons stated above, it will be difficult to gain the full benefit of genetics and genomics without deep clinical decision support. As we dive deeper into the implementation of this clinical decision support, various challenges emerge. Some are technical, while others relate to business issues. But perhaps most challenging are those surrounding legal responsibility.

As discoveries are made relative to genetic variants that have been found in our patients, whose responsibility is it to bring that new knowledge into the care delivery process and make sure it is acted upon? In today's world, the answer appears to be "no one's." This is not to say these types of updates do not happen. Our geneticists have worked overtime to notify physicians when particularly significant new information is identified. However, when this happens in today's world it takes the form of an unreimbursed activity that involves clinicians going above and beyond the call of duty.

We are currently building infrastructure that will simplify the dissemination of these "knowledge events." One concern we have encountered is that the act of developing this infrastructure will create legal liabilities on clinicians that do not exist today. Once we have completed an infrastructure capable of disseminating knowledge events, will geneticists or practice leaders become liable for not identifying new genetic knowledge that could be useful to patients? This is a real concern because of the number of places such knowledge could be reported or stored. It would be nearly impossible to develop a standard operating procedure (SOP) that would be guaranteed to identify relevant new knowledge in a timely fashion. Even if such an SOP could be developed, it would be extremely expensive to maintain. IT can help by integrating with external databases but it cannot provide a complete solution to the problem. This leads to the question "if we can't do everything, can we do anything?" Government could help us make the

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

answer to this question "yes." Everyone benefits from the creation of genetic-based clinical decision support infrastructure. Even if it does not result in comprehensive knowledge updates to patient genetic profiles, it can provide clinicians with vastly improved information and guidance. We need government, perhaps through the Institute of Medicine, to develop guidelines that, if followed, would significantly limit any liability exposure associated with installing these systems.

On the flip side, it is also important that the government act to ensure our ability to leverage all information available to improve the health of our patients. This includes all information that might be known about the variants identified in our patients. Ideally our systems should not be responsible for tracking intellectual property restrictions on the use of particular variant information relative to patient care. We should be empowered to use this information in any way that would optimize care.

Providing a Basis for Interacting with Personal Health Records

Patients are likely to desire more access to their clinical information over time, particularly as the amount of genetic information in their profiles grows. Personal Health Records (PHRs) represent a means of providing this access. PHRs are similar to EMRs except they are managed by an individual instead of an institution. Therefore they are portable. They also provide the patient with access to all the information contained within them. There is a strong argument that PHRs could lead to fundamental improvements in the way health care is structured in the United States. However, PHRs will work only if providers develop interfaces to feed data to them. Government could assist by developing regulations to govern this type of transfer and clarify the circumstances under which it should take place.

Ideally, PHRs would assist in educating patients about the implications of their genetic variants. It is hoped that PHR providers will take care to do so in a robust and responsible manner so the issues noted in the first section do not materialize.

A Unified-Idealized Scenario

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

The following scenario is meant to provide a vision for how our enterprise IT infrastructure supporting genetic and genomic based personalized medicine might function once it is fully developed.

Patient Visit

Mr. Smith schedules an annual physical with his physician, Dr Jones. In advance of the visit, Mr. Smith is asked to log into our patient web portal and review the background information in his medical record. He notes that since his last visit, one of his first degree relatives died suddenly from a heart-related complication. This alerts Dr. Jones that Mr. Smith may be a candidate for an HCM molecular diagnostic test. When Mr. Smith arrives for his visit, Dr. Jones discusses the pros and cons of HCM testing. They decide to order the test. When Mr. Smith's blood is drawn, he is asked whether he would consent to participate in research studies. He indicates he is willing to participate in studies focused on heart disease but no other types of studies. This information is recorded in the CTS. Mr. Smith's blood is sent to our Laboratory for Molecular Medicine. Mr. Smith had genetic tests performed previously and is concerned about a report he recently read online that one of the variants identified in his DNA was found in many colon cancer patients. Dr. Jones queries GeneInsight by clicking on the variant in the CDR. Dr. Jones sees that this variant is only implicated in colon cancer when combined with other variants that she can confirm Mr. Smith does not have. Dr. Jones explains this to Mr. Smith and he is relieved.

The Diagnostic Testing Laboratory

Fortunately for Mr. Smith, our Laboratory for Molecular Medicine has recently installed and was able to quickly migrate to new testing technology. This technology lowered the cost to the point where Mr. Smith's insurance company decided it was cost-effective for them to reimburse for the test. The sample is logged into GIGPAD. GIGPAD tracks the sample through the testing processes, which identifies several variants, including OLDGENE 872G>C. These variants are sent to a geneticist who generates a report. In generating this report, the geneticist consults GeneInsight. GeneInsight contains all of the annotations previously generated by the Laboratory for Molecular Medicine plus a rich set of annotations pulled from external databases identified by our geneticists as credible sources of information. GVIE generates a draft report that is approved by the geneticist. Based on

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

the best information available, all of Mr. Smith's variants are classified as "presumed benign." The test results are sent to the electronic medical record. Dr. Jones informs Mr. Smith his test was negative.

The Research Process

Three months later, RelNet identifies that patients reporting a family history of sudden death are more likely than the general population to have OLDGENE 872G>C. A clinical researcher, Dr. Adams, reviews this result and decides the finding is worthy of further investigation. Dr. Adams runs an RPDR query that identifies that while there are a large number of deceased patients with this variant, it is also found in a significant number of older, healthy patients. Dr. Adams leverages the SPIN network to identify banked tissue samples from patients who succumbed to sudden death. Dr. Adams receives permission from both the bank owners and her IRB to assay the samples. Dr. Adams also receives approval to establish prospective criteria to capture samples from patients with this variant that flow through pathology laboratories that would otherwise be discarded. Dr. Adams uses GIGPAD to order genotyping analysis on the samples collected from the deceased patients and expression profiling analysis on all the samples collected. The expression analysis identifies a gene that is under-expressed in all patients with the variant who succumbed to sudden death, but only some of the living patients with the variant. Dr. Adams leverages an I2B2 hive cell to identify promoter genes that regulate the under-expressed gene. She then leverages GIGPAD to order sequencing analysis focused on these promoters. She identifies a protective variant, NEWGENE 9926G>A, that is found in all the elderly patients living with the initial variant but none of the patients who succumbed to sudden death. Based on her finding and after consulting with the appropriate clinical governance bodies, she updates the interpretation of OLDGENE 872G>C to "pathogenic." She also adds another entry into GeneInsight for NEWGENE 9926G>A, which she classifies as "protective" relative to OLDGENE 872G>C.

The Laboratory

The finding of Dr. Adams is considered significant enough to warrant development of a new molecular diagnostic test. Because all the research work was done in the GIGPAD context, a significant amount of information about test configuration can be quickly migrated to the clinical environment.

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

A new test is quickly launched. Information about how this test improves the sensitivity of older HCM tests is recorded in the GeneInsight.

The Clinical Decision Support System (CDSS)

Our CDSS identifies that a new, more sensitive HCM test is available and is particularly relevant to patients who have OLDGENE 872G>C. Because Mr. Smith has this variant, Dr. Jones is sent an alert indicating that Mr. Smith is a candidate for the updated test. Dr. Jones speaks with Mr. Smith about obtaining the new test, which he decides to do. When Mr. Smith's blood is drawn, the phlebotomist reviews the CTS record indicating Mr. Smith's willingness to participate in heart studies. Mr. Smith is feeling anxious about his medical condition and decides he would rather not be contacted about research studies. This information is entered into the CTS. The test is completed and indicates that Mr. Smith does not have the protective NEWGENE 9926G>A variant. The CDSS indicates that an implantable cardioverter defibrillator (ICD) should be considered for patients with Mr. Smith's genetic profile. Dr. Jones discusses the situation with Mr. Smith and he decides he would like an ICD. When Mr. Smith is preparing for surgery, his genetic profile is leveraged in drug dosage calculator and protocol selection routines to ensure that he receives the optimal treatment for him. He does well in surgery. Six months later he has a major cardiac event and the ICD saves his life.

Conclusion

Genetic and genomic knowledge is revolutionizing our understanding of human health and disease, and incorporation of this knowledge into clinical practice is expected to bring in the new era of personalized medicine. Incorporation of genetic and genomic information into clinical decision making provides a number of IT challenges. The Partners IT group, the Harvard-Partners Center for Genetics and Genomics, and the Hewlett Packard Corporation have developed and implemented a solution to these IT problems to help patients treated at our hospitals in the Boston area. It is possible that these systems might serve as models for other institutions or at a national level.

We very much appreciate the request of the federal Department of Health and Human Services for information related to how IT can support the

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved

advance of personalized medicine. This is a highly dynamic area that has great potential for improving patient care. At the same time, the burgeoning field of personalized medicine would benefit from government assistance in making sure it grows in an efficient manner that maximizes the benefit to the patient. Thank you for this opportunity to share our views on this subject.

References

1. Collins F. The Genome Project: Changing the Paradigm of Research in Life Sciences. Paper presented at: IOM Annual Meeting, 2002.

2. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol.* Dec 2001;19(12):491-496.

3. Trotta R, Donati MB, Iacoviello L. Trends in pharmacogenomics of drugs acting on hypertension. *Pharmacological Research.* 2004;49(4):351-356.

4. Weiss ST, Litonjua AA, Lange C, et al. Overview of the pharmacogenetics of asthma treatment. *Pharmacogenomics J.* 2006.

5. Burrill GS. *Biotech 2006 Life Sciences: A Changing Prescription.* San Francisco, CA: Burrill and Company; March 24, 2006 2006.

6. Middleton KR, Hing E. National Hospital Ambulatory Medical Care Survey: 2004 outpatient department summary. *Adv Data.* Jun 23 2006(373):1-27.

7. Murphy, S.N., Chueh, H (2002). A Security Architecture for Query Tools Used to Access Large Biomedical Databases. AMIA, Fall Symposium 2002, pages 552-556

Copyright 2007 – Harvard Medical School – Partners Healthcare Center for Genetics and Genomics – All rights Reserved