

**Information Technology for
Genetic and Genomic Based Personalized Medicine**

Submitted

April 23, 2008

By

**The Harvard Medical School – Partners HealthCare
Center for Genetics and Genomics**

in Partnership with

The Partners HealthCare Information Systems Department

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 increased use of genetics and genomics in medicine accelerates, substantial new pressures will be placed on our health care delivery system. Some 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 over four and a half years to develop and implement strategies to meet the needs of personalized medicine. The collaboration focuses on enhancing enterprise information technology infrastructure that supports both research and clinical environments. This document describes the infrastructure we have deployed as well as additional enhancements we believe are necessary to prepare us for wider adoption of genetic and genomic techniques in medicine. We will also suggest ways that governments may be able to assist in the development of this type of critically needed infrastructure.

The Nature of Genetic Based Diagnostic Tests

Molecular diagnostics, also known as genetic or genomic based diagnostics, provide the bridge that enables physicians to bring genetic knowledge to routine patient care. Physicians use the information these tests generate for assessing disease risk, for diagnosis, for prognosis and for making specific treatment decisions. At our Center we offer molecular diagnostics to:

- Diagnose single gene disorders. For example, determine whether patients have genetic variants that cause certain types of cardiac disease, some of which are severe enough to result in sudden death if left untreated, and to assess the risk of their relatives for these disorders.
- Identify whether a patient’s hearing loss is caused by genetic variants that are also correlated with other serious medical conditions.
- Determine whether non-small cell lung cancers have genetic variants that correlate with either Tyrosine Kinase Inhibitor (TKI) efficacy or resistance.
- 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).

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 in one gene or many genes and determine all the variations present in those sequences. Chip-based technologies have recently broken through into the clinical arena and have made it possible to survey increasingly large segments of DNA in a cost-effective manner. Several national governments are funding research and many commercial entities are investing significant capital to reduce the cost of sequencing a person’s entire genome to approximately \$1,000. While we are a few years away from reaching this goal, new technologies will continue to drive down the cost of sequencing to costs that may be even less than \$1,000/genome.

The continuous reduction in DNA sequencing costs will significantly affect how genetics is leveraged in the clinic. At present the cost of DNA sequencing is a major barrier to increased adoption of molecular diagnostic testing. As this barrier is reduced, we believe the amount of patient DNA being sequenced will increase. We expect the number of variants identified in the patient population to grow continuously until it becomes feasible to cost effectively sequence all of the nearly 3 billion base pairs of a patient’s genome. Current estimates place the number of variants that such a test would yield at 4-5 million per person. (Levy et al. 2007) The data generated by such a

whole genome sequencing test would be good for a person's lifetime. Sequencing would only need to be redone in the case of cancer or other disorders where somatic changes are important and for infectious agents.

We are rapidly approaching the day where we will be able to determine the precise DNA variations present in each patient. However, the process of determining the implications of each of these variants, let alone the implications of each combination of variants, will take longer. Our knowledge of the impact of genetic variation is constantly expanding and this knowledge expansion is likely to continue for many years to come. Ideally, clinicians would take into account the most up to date discoveries on every variant discovered in each patient as they prescribe care, but doing so is clearly beyond the capacity of the human mind. To reach this goal, clinicians will require far more extensive IT infrastructure than what exists today. Infrastructure of this magnitude takes time to build. For the past five years the Harvard Partners Center for Genetics and Genomics, the Partners HealthCare Information Systems department, and Hewlett Packard have been developing such an infrastructure. The applications we have built together are supporting the use of genetics in the clinical environment; however, much work remains to be done to provide the depth of support clinicians will ultimately need.

The Partners HealthCare Genetics IT Infrastructure

To be truly effective, IT infrastructure that helps manage genetic information must integrate laboratories, geneticists, the Electronic Health Record (EHR) and automated clinical decision support engines. Infrastructure of this scope must be built incrementally. In our institution, we began by constructing a platform to support the laboratories that generate genetic and genomic data. Next we built infrastructure that supports professional genetic experts and other healthcare professionals including genetic counselors.

Then we integrated this infrastructure with the Partners HealthCare EHR. Finally we began the work of creating genetics based clinical decision support (CDS) functionality. We are now deepening our support for genetics in the EHR to enable broader genetics based clinical decision support. (Figure 1) As we do this, we are encountering challenges that cannot be solved by an individual organization acting independently. Therefore, we are working to establish networked infrastructure that will be needed to fully support personalized medicine.



Figure 1: The Process of Standing up IT Support for the use of Genetics in Medicine

Supporting the Laboratory: The Gateway for Genomics-Proteomics Applications and Data (GIGPAD)

Almost all genetic and genomic data are generated in laboratories by complex machinery. 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, reducing errors, 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. Maintaining multiple LIMS within an enterprise is both challenging and expensive. We have found that creating an Enterprise LIMS Superstructure can help address these problems. In our environment we created a system called the Gateway for Integrated-Genomic Proteomic Applications and Data (GIGPAD) to serve this function. GIGPAD serves as an umbrella over the individual LIMS in the environment and integrates them together. The system exposes unified user and system interfaces to the rest of the enterprise.

GIGPAD has been live in our environment since April of 2004. As of April 18, 2008 there were 885 registered users of the system and 947,666 data files under management. GIGPAD currently provides support for the initial phase of the Molecular Diagnostic testing process. This includes all steps up until the point that we determine what genetic variations are present in the stretches of DNA that are sequenced. At this point GIGPAD forwards this information for interpretation (Figure 2). GIGPAD is designed to handle DNA based, RNA based or protein based testing efforts.

Supporting the Geneticists: GeneInsight and the 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.

We have constructed two tools to support the report generation process in our environment: GeneInsight and the Genomic Variant Interpretation Engine (GVIE). Because our understanding of the clinical implications of particular variants can change over time, it is important to have a database that tracks current knowledge relative to individual variants. We use GeneInsight to perform this function. Keeping this type of database current is extremely challenging. There are numerous heterogeneous research databases that contain information about genetic variants but very few clinically validated data sources. Geneticists and genetic counselors must review these research sources to formulate clinical interpretations. 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 GVIE. GVIE is a reporting tool that is interfaced to GIGPAD.

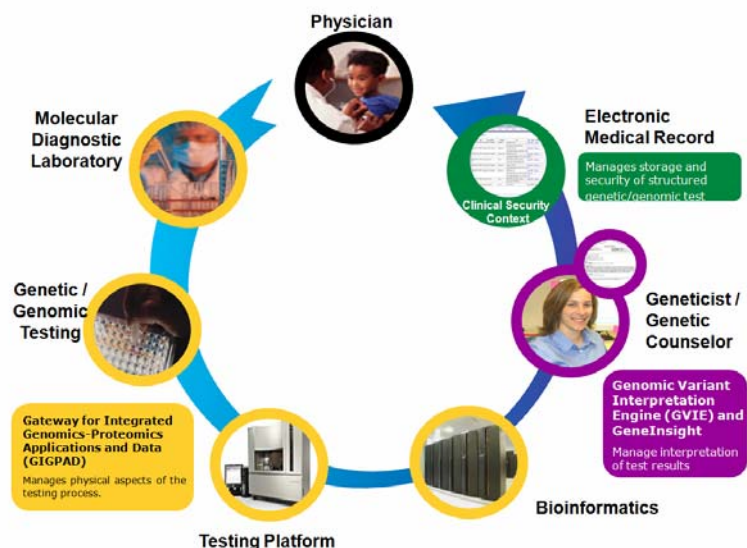


Figure 2: Applications Used to Support Molecular Diagnostics within Partners HealthCare. As variants are identified in patients, they are passed to GVIE which then looks up the information stored in GeneInsight on those particular variants. GVIE then produces a draft interpretive report which a geneticist and/or

genetic councilor reviews. During this review process, they are shown statistics related to the variant's frequency and given the ability to review previous cases where the variant was identified. Geneticists and genetic councilors have the option of modifying these reports. We track which reports are modified. This provides us with a metric for assessing the maturity of each part of GeneInsight.

As a result of this process, geneticists can maintain the data in 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 amount of data contained in GeneInsight has grown over time and we are now evaluating additional uses for this information in the clinical environment. However, as we will describe later, we need to find ways to dramatically increase the depth and breadth of the data in GeneInsight if it is to solve our core genetics related knowledge management needs.

Supporting Front Line Clinicians: Electronic Health Record (EHR) and Clinical Decision Support (CDS) System Integration

Molecular diagnostic reports are ideally delivered to the clinician through an EHR. Doing so ensures that genetic test results are stored in an organized manner and are consistently accessible to authorized clinicians. It also opens up the possibility of leveraging automated CDS systems to proactively assist clinicians in the use of this information. We created a specially secured area in our EHR where we maintain patient genetic profiles. GVIE is interfaced to this part of our EHR in a manner that allows us to transfer reports in both human readable and highly structured electronic formats. The structured report format is designed to be read by CDS algorithms.

As the number of variants stored in patient genetic profiles increases, it will become increasingly difficult for clinicians to review these profiles during the care delivery process. Properly applying the information in these profiles will be even more challenging. Clinicians will need to rely on CDS functionality to surface relevant genetic information at the appropriate times. This functionality is required for genetics aware personalized medicine to reach its potential, but it will be very difficult to build. We have taken an initial step within Partners by establishing our first genetics aware clinical decision support rule. This rule alerts physicians if they order Iressa or Tarceva for a patient who has a genetic mutation associated with resistance to these drugs. We are now constructing an infrastructure that will make it easier to setup additional rules in the future. In doing so, we have learned that establishing a truly robust genetics aware CDS capability will require us to develop a large number of interfaces with other institutions.

Linking it all Together: Establishing the Data and Knowledge Flows Needed to Drive Genetics Aware CDS

In our environment we have established a flow that links together GIGPAD, GVIE, GeneInsight and our EHR. When a Partners HealthCare patient is tested in our Laboratory for Molecular Medicine (LMM), the results flow into our EHR in structured form. Up to date knowledge about the implications of any variations found by the LMM is maintained in GeneInsight. When we test our own patients, we have both the knowledge and data resources required to construct genetics based CDS.

Genetic tests are performed by many different laboratories throughout the world. Many of the genetic tests performed on our patients are performed by external laboratories. Similarly, our Diagnostics laboratory often tests patients for other providers. In both of these cases, interfaces do not exist to transfer the variants identified in electronic form. Therefore, the structured genetic data is not ultimately represented in an EHR. Without these data, CDS is impossible. In addition to this data transfer problem, there is also a knowledge maintenance issue. CDS systems should ideally be linked to databases like GeneInsight so that the rules they run can take into account the most up to date genetic knowledge. The information in GeneInsight is continuously updated as our knowledge of

the significance of individual variants improves. The problem lies in the fact that because GeneInsight is curated by our geneticists, it only contains information on the genes covered by our LMM tests. No single institution could ever hope to employ enough geneticists to maintain up to date information on all of the variations that could be identified in its patient population.

We believe that inter-institutional networks must be established to overcome these issues. This involves both standards development and infrastructure construction. Our institution is investing heavily in both areas. A member of the HPCGG IT team serves as one of the co-chairs of the HL7 Clinical Genomics Special Interest Group. We have developed and contributed internal message formats to HL7 and worked with them to develop a standard model to transfer genetic test results. We have also worked extensively with the leadership of LOINC to establish appropriate coding schemes for genetic tests. We also interact with government institutions focused on supporting the development of standards for personalized medicine data.

In addition to developing standards, we are also investing in two infrastructure projects designed to facilitate the exchange of genetic data and knowledge. First, we are exploring the possibility of making a service, based on GeneInsight, available to external institutions. This would allow external laboratories, and ultimately external providers, to take advantage of the functionality GeneInsight offers. It could also simplify the process of sharing (non-patient specific) knowledge across institutions. This will be a non-trivial process. Each institution's knowledgebase will need to be appropriately secured so that knowledge transfers occur only with their authorization. However, we believe it is possible to setup an environment where each institution could be incented to release their knowledge for clinical use.

We are also working on another infrastructure called VariantWire. VariantWire will serve as a hub that will enable the secure transfer of genetic test results. Any institution that connects to VariantWire will be able to send genetic test results to any other connected institution. We are building validation functionality into the hub that will enable us to enforce standards by rejecting any non-conforming messages. VariantWire is intended to address the "many-to-many" problem. In theory, each provider must establish an interface to each genetic testing laboratory if they want to develop and maintain comprehensive patient genetic profiles in their EHR. The cost of these interfaces makes this impractical. VariantWire will address this problem by allowing providers and laboratories to connect to multiple institutions through a single interface.

Steps Governments can take to Accelerate Development and Adoption of Clinical Genomics 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. 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, government established exceptions that are a response to a significant public interest.

Governments can also play a role in establishing an economic environment that promotes the growth of personalized medicine. It is particularly important to think about this in the context of appropriate regulatory infrastructure that promotes use of genetic/genomic information in determining drug safety and efficacy. The expenses associated with genetic tests are moving from the physical testing process to the interpretation process. The healthcare system must acknowledge this shift and provide mechanisms for independently compensating the process of generating interpretations. Developing the IT support for the practice of personalized medicine in general will become easier if these compensations mechanisms are put into place.

Reference

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, et al. (2007) The diploid genome sequence of an individual human. PLoS Biol 5(10): e254. doi:10.1371/journal.pbio.0050254