For reprint orders, please contact:
reprints@futuremedicine.com

PERSPECTIVE

Drug safety assurance through clinical genotyping: near-term considerations for a system-wide implementation of personalized medicine

Michael D Kane[1,2†], John A Springer[1] & Jon E Sprague[3]

†Author for correspondence
[1]Purdue University, Department of Computer and Information Technology, West Lafayette, IN 47907-2021, USA
Tel.: +1 765 494 2564; Fax: +1 765 496 1212; E-mail: mdkane@purdue.edu
[2]Purdue University, Bindley Bioscience Center at Discovery Park, West Lafayette, IN, USA
[3]Ohio Northern University, Raabe College of Pharmacy, Ada, OH, USA

The rationale and overall system-wide behavior of a clinical genotyping information system (both DNA analysis and data management) requires a near-term, scalable approach, which is emerging in the focused implementation of pharmacogenomics and drug safety assurance. The challenges to implementing a successful clinical genotyping system are described, as are how the benefits of a focused, near-term system for drug safety assessment and assurance overcome the logistical and operational challenges that perpetually hinder the development of a societal-scale clinical genotyping system. This rationale is based on the premise that a focused application domain for clinical genotyping, specifically drug safety assurance, provides a transition paradigm for both professionals and consumers of healthcare, thereby facilitating the movement of genotyping from bench to bedside and paving the way for the adoption of prognostic and diagnostic applications in clinical genomics.

Keywords: clinical genotyping, data management, electronic health record, personalized medicine, pharmacogenetics, SNP

The utilization of a patient's genetic data to aid diagnostic and prognostic healthcare represents the ultimate achievement of 50 years of genomic research. The technology to recognize this vision has emerged, and continues to evolve. In the future, patient-specific genomic data will be derived before birth and will include an exhaustive sampling of genomic information. These genetic data will be periodically updated throughout a patient's lifetime on a tissue-specific basis in order to screen for genetic changes conferring age-related diseases. The patient's genotypic data will further be integrated, with dedicated databases/warehouses harboring genetically linked health and adverse drug response risk, which will be utilized at the point of care for patient-specific therapeutic interventions. However, the path to this future in genomic-based healthcare is obscured by several independent factors that must be recognized and overcome to fully exploit genomic content in human healthcare. These hindrances are not just limited to those associated with pharmacogenomics [1,2], but with any system-scale utilization of human genetic data in healthcare. The authors have identified the following categorical hindrances to a societal-scale implementation of clinical genomics:

- High-throughput DNA analysis technology: costs, data standards and future technologies;
- Information management: access, security and system structures;
- Genomics and genetics education: physicians, pharmacists, nurses and consumers;
- Point-of-care utilization of genomics: physician's office, hospital, pharmacy and consumer;
- Capitalism and pharmaceuticals: risks and returns on investment in genomic-based laboratories and information systems;
- Electronic health-record (EHR) management and utilization;
- Translational research: establishing linkages between allelic information and healthcare outcomes.

Many of these issues are addressed (directly or indirectly) in this manuscript through near-term operational system solutions. Ultimately, the authors attempt to lay the groundwork for implementing a near-term system for clinical genotyping, including:

- The information management system and data standards required for system implementation
- The interface between DNA analysis biotechnology and a clinical genotyping information system
- The management of costs and opportunities to insure the success of a clinical genotyping system
- The educational impact of the postgenome era on pharmacy students and other professions within healthcare

High-throughput DNA analysis technology: costs, data standards & future technologies

There are numerous analytical methods for DNA analysis that support both SNP discovery

and SNP detection (SNP discovery and detection represent distinct analytical challenges that are not described in this manuscript), and competition within the biotechnology industry continues to advance these capabilities from all relevant perspectives (cost, throughput, data quality, ease-of-use and so on). Importantly, regulatory agencies, such as the US FDA and European Medicines Agency, have supported the emergence of pharmacogenetic screening through the publication of testing and technology guidelines [101]. Yet at the heart of a large-scale clinical genomics implementation is an information management system that can accommodate many different analysis methods (including new biotechnologies that will emerge in the future) through the development of a group of scalable data standards for genomic information. Although significant advances in biotechnology are occurring, the data standards for sharing genomic data will precede genotyping in the clinic. Importantly, a short turn-around time for a genetic screening test will be central to establishing a value proposition for this emerging facet of healthcare. Ideally, a stand-alone instrument will be utilized by non-technical staff (i.e., clinical staff) to screen clinically relevant SNPs at the point-of-care facility using a patient's buccal swab sample, and results will be presented in minutes (not days). As an example, SNP screening at the initiation of warfarin therapy (i.e., screening for allelic variants in *CYP2C9* and *VKORC1*) will be very useful in immediately identifying patients who are warfarin sensitive (i.e., poor metabolizers of CYP2C9) or warfarin resistant (i.e., altered warfarin binding in VKORC1), but this will only be therapeutically useful (and cost effective) if results from the SNP screen are available before or very early in the dosing regimen, and prior to International Normalized Ratio results from the patient. Thus, the verification of positive impact on healthcare costs through a pharmacoeconomic assessment will hinge upon the turn-around time of the technologies employed for SNP screening and the utility of the results to reduce the overall incidence of adverse drug responses. Ultimately, this will influence the acceptable cost of screening, whether this is US$20 or US$200 per test, in the near future. This model of onsite gene screening could be positioned to significantly reduce the costs of SNP screening, which currently depend upon basic research instrumentation. Since most of the patients tested for pharmacogenetic markers will be identified as 'normal', the onsite SNP screening system need only be capable of rigorously identifying the normal state, and any sample that harbors evidence of a variant allele will be subsequently assayed by a more specialized technology in a regional genetic screening laboratory.

Information management: access, security & system structures

Included in this manuscript is the rationale for an information system that categorically separates SNP data relevant to drug safety from SNP data relevant to general health outcomes. The authors feel strongly that this distinction is key to the implementation of a successful clinical genotyping system. By categorically separating SNPs relevant to drug safety from SNPs linked to other health outcomes and SNPs with no known linkages (it is recognized that there is some small overlap in this distinction), consumers can:

- Understand how their own genomic data is being utilized and gain trust in these systems
- Indicate how their own genomic data is managed and who can gain access to these categorical datasets
- Provide a rationale for security that is dependent upon the category of the data

For example, drug safety data may be more easily accessed by worldwide healthcare institutions and pharmacies, since these data may be needed in an emergency for an injured traveller. By contrast, other SNP categories are stored much more securely and are not shared across institutions. This concept assumes that consumers will be able to control access to their genotypic information and that SNPs inherent to drug safety are far less likely to serve (or be abused) as indicators of general health for an individual.

Genomics & genetics education: physicians, pharmacists, nurses & consumers

Given the very recent advances in human genomic knowledge and biotechnology methods, it is not feasible to assume that physicians, pharmacists, nurses and other professionals within the healthcare industry harbor sufficient knowledge to translate raw genomic data to information relevant to health outcomes [3]. This being said, the authors have made two conclusions regarding the near-term future of clinical genomics. First, essentially all genomic data will

be filtered into categorical definitions and the known (or potential) impact of a given SNP will be presented to the healthcare professional (described in Table 1). For example, if a patient is prescribed a drug where an adverse response has been associated with one or more specific genotypes, the patient's EHR will simply indicate that the patient is "at risk for an adverse response due to genomic information" and make a recommendation to choose an alternative drug (and provide an alternative drug if one is available) and/or reduce the dose of the drug. This specific example, as well as many others, has recently been demonstrated in the clinical literature. Drugs are metabolized endogenously by a series of enzymes collectively referred to as the CYP system. These enzymes are further characterized into subgroups, named CYP1A1, CYP2D6 and so on. Meur *et al.* demonstrated that the metabolic activity and oral clearance of the immunosuppressant sirolimus is significantly decreased in patients with the *CYP3A5*3* SNP and further suggested that prior dose adjustments should be made in patients with this SNP [4]. However, the technology for routinely implementing such a dose adjustment does not currently exist. The initial commercially viable implementation of a clinical genomics system will clearly involve drug safety issues and be administered through pharmacy prescription systems. This initial implementation of a pharmacogenomic system will utilize SNPs that have an established link to drug safety outcomes [5] and therefore can include information-based guidance to patients harboring SNPs relevant to drug safety (i.e., decision support for both the physician and pharmacist), and can exploit a prescription/dispensing system that is already guided by an information system and inherently does not involve SNPs poorly linked to disease risk and/or does not provide insight on how a physician or pharmacist should alter treatment. Furthermore, this near-term implementation will provide a cultural shift in pharmaceutical drug development whereby new drug indications can require genomic screening to ensure safety and efficacy, ultimately involving clinical drug development (Phase I–IV) to be limited to patients with specific SNP genotypes in order to increase the overall safety and efficacy of new drug entities.

Point-of-care utilization of genomics: physician's office, hospital, pharmacy & consumer

This issue continues the rationale for a near-term implementation of clinical genomics in drug safety by allowing pharmacists to be the proprietors of genomic information and exploiting the interconnectivity of pharmacy information systems to allow access to patient genomic information across the country. It is likely that other gene products relevant to drug absorption, distribution, metabolism and elimination, such as drug transporters, will quickly emerge as valid allelic markers for estimating drug and/or dose efficacy and safety. As many more SNPs are ultimately derived for each patient, a more secure healthcare information system will include SNPs relevant to disease predisposition as they are established through translational research. As discussed later, the translational research that involves linking known SNPs to healthcare outcomes will be facilitated through the use of the near-term implementation genotyping system.

Capitalism & pharmaceuticals: risks & returns on investment in genomic-based laboratories & information systems

The implementation of a drug safety clinical genomic system provides the best overall return on investment for the healthcare community in the near-term. The established prescription/dispensing system can accommodate the near-term drug safety assurance objectives, where the

Table 1. Information hierarchy within a clinical genotyping information system.

Conceptual perspective	Data schema perspective	Access and privacy perspective	Bioinformatics perspective
Data	Physical data	Laboratory	Raw data (DNA)
Information	Conceptual data	Bioinformatics Statistics Epidemiology	DNA sequence Bioinformatics Genomics and ontology
Knowledge	User view	Healthcare Data management Applications	Protein sequence Enzyme biochemistry Receptor biochemistry
Comprehension	Consumer view	Patient	Physiological effect Impact on healthcare

drug–drug interactions that are currently routinely screened in this system are augmented with a gene–drug interactions' screen under a similar drug safety objective. Ultimately, the adoption of this system will lead to a shift in pharmaceutical drug development where new drug indications include specific genomic screens to ensure overall safety and efficacy of new drug entities.

Electronic health-record management & utilization

The utilization of an EHR is a very new concept in healthcare, and although the benefits of an information management system that utilizes EHRs are essentially well recognized, the implementation of EHR systems has not been widely successful to date. Therefore, it may be too ambitious to assume that the inclusion of clinical genomic data in current EHR systems will be inherently successful. The authors feel that the overall usefulness and impact of genotypic information in the clinic (from both the consumer and healthcare provider perspective) must precede a widespread system implementation. This further rationalizes the system described below, where SNPs relevant to drug safety, as utilized in the pharmacy (and pharmaceutical industry), represents the ideal introduction of genotypic information in our healthcare system.

Translational research: establishing linkages between allelic information & healthcare outcomes

This logistical barrier to the overall impact of genotypic information in the clinic involves a disparity between discovering (or uncovering) linkages between known SNPs and human health, which requires a large collection of known SNPs from a wide variety of patients (including their health records within one or more data standards), and a method regarding how to rationalize the collection of known SNPs from a wide variety of patients. In other words, statistically significant linkages between known SNPs and health outcomes can only be achieved if a large collection of SNPs from normal and 'diseased' patients is available for data mining. Furthermore, this requires that the disease-relevant information and other metadata types be available within data standard formats to allow for data mining, which is the fundamental structure of an EHR. The authors feel strongly that the near-term drug safety system that integrates known SNPs with prescription drug indications will facilitate the acquisition of many other known SNPs that are not relevant to drug safety for the purposes of epidemiological research. In other words, patients undergoing genotyping for drug safety will have the option (ideally with incentives) to be genotyped for thousands of other known SNPs within their own genome to facilitate health outcomes research, ultimately to benefit themselves and society. As discussed below, this will involve an anonymous contribution of SNP and EHR data to a specialized data-management system dedicated to identifying SNP-based risk assessment through the discovery of statistically significant linkages to other health outcomes, such as diabetes, cancer, mental disorders, age-related disorders and so on. This concept gives rise to an oversight committee that governs data mining and statistical methods to establish 'accepted' links between SNPs and health outcomes, and 'approves' new linkages as they are discovered, proven and published.

Data management can be viewed from two perspectives, where the overall concept of 'informational hierarchy' is used to describe both data concepts and data schemas (moving from left to right in Table 1), which then define levels of information access (privacy and security) and levels of bioinformatics knowledge (raw biotechnology data to DNA sequence to protein sequence to physiological effect). This informational hierarchy (Table 1) is also organized vertically (top to bottom) to depict data transformations from raw data (biotechnology and DNA-analysis data) into usable information (bioinformatics) and comprehensible knowledge (impact on human health).

Clinical genotyping for drug safety

System-wide operations

Patient-controlled access

The ethical concerns to genotyping in the clinic, which are also applicable to EHRs in general, are essentially privacy and security. The benefits of incorporating genotyping (genetic information) in therapeutics and medicine are questioned when the risk of 'information abuse' is considered. For example, a patient may be unwilling to utilize the benefits of genotyping if they fear that their employer and/or insurance provider can utilize the same information to (accurately or inaccurately) predict the patient's future health status. This dilemma involves both societal and genetic components. At the genetic level, the validity of extrapolative health assessment based

solely on genotypic data has not been broadly established, and is limited to a few known genetic diseases. Therefore, any long-term claims to health status for the majority of the population would be invalid at this point in time. Yet, it should be noted that the risk of adverse drug response based on known SNPs in drug-metabolizing enzymes has been established (see Table 2), and represents the near-term benefit to clinical genotyping.

Furthermore, it is important to note that the use of the term SNP herein includes nucleotide base substitutions and single-base deletions/substitutions within the human genome. In addition, knowledge of this predisposition does not represent association with other health risks. Thus, knowledge of the risk of adverse drug response is a benefit to the patient, employer and insurance provider, since overall healthcare costs would be minimized by avoiding adverse drug reactions. Allowing the patient to control external access to their genotypic data within this categorical distinction (e.g., 'adverse drug response risk' data access = yes; 'general health risk' data access = no) will positively contribute to the adoption and success of genotyping in the clinic (see Figure 1), which is a natural artifact of utilizing the hierarchy described in Table 1.

Human genotypic database

The genotypic data derived from large numbers of people and patients will be crucial in establishing new genetic links to health risk [6]. In this paradigm, it is assumed that a large cohort of patients will be genotyped across thousands of known SNPs that include SNPs that have established links to the risk of adverse drug responses and disease, as well as SNPs that currently have no known association with human health outcomes. The discovery of one or more SNPs that is associated with a specific phenotype or disease risk requires that a large Human Genotypic Database (HGD) is derived from individual genotypic records, which includes all other aspects of their health records. For example, the discovery of SNPs that are linked with cardiovascular disease involves a statistical comparison of SNPs between a large group of patients experiencing cardiovascular disease and a large control (disease-free) group. In practice, this involves the derivation of a HGD where the patient identifiers have been removed (achieving privacy through anonymity). These include genotypic and overall health information for each person, which is a natural artifact of utilizing the hierarchy described in Table 1.

Risk database

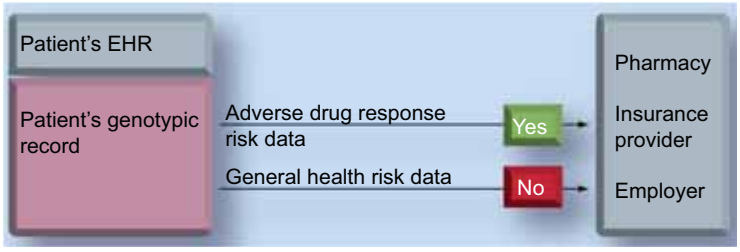
The incorporation of health 'risk' data, which is the known risk associated with each SNP position, into a patient's genotypic record must be temporary and periodically updated to reflect new discoveries and linkages. This dynamic component to the EHR reflects the fact that future

Table 2. Representative drug-metabolizing enzymes associated with adverse drug response‡.

CYP family	Allele	Nucleotide change	Enzyme activity change	Impact on AUC	Common drugs affected	Ref.
1A2	*CYP1A2*1C*	-3860 G>C	Decreases	Increases	Aminophylline Betaxolol Caffeine Flutamide Propranolol	[10]
2C9	*CYP2C9*3A*	1075 A>C	Decreases	Increases	Amiodarone Carvedilol Fluoxetine Glimepiride Warfarin	[11]
3A4	*CYP3A4*18A*	878 T>C	Increases	Decreases	Atorvastatin Carbamazepine Clarithromycin Diltiazem Losartan	[12]

‡This is not an exhaustive representation of known SNPs relevant to drug metabolism and safety, and only three representative CYP families, a single SNP example, and five potential drugs affected are presented in this table for conservation of space.
AUC: Area under the curve of a drug's concentration in human plasma over time.

Figure 1. The patient controls outside access to genotypic data based on how the data is used (categorically).



In this example, the patient has allowed access to SNP data corresponding to adverse drug response risk, yet prohibited access to SNP data known (or unknown) to be relevant to overall disease and general health risk. Note that the patient's genotypic record is regularly updated with new genetic data from the risk database, and the categorical limited access to the health record will allow access to new adverse drug response risk data.
EHR: Electronic health record.

discoveries may link known SNPs to one (or more) health outcomes; in the absence of an updatable risk component, a patient's genotypic record will become outdated and, thus, underutilized. For example, a patient may have data on a specific genotype (SNP or set of SNPs in a specific genomic location) that, to date, has been considered benign and represents no known risk, yet new research findings have determined that the SNP constitutes some level of health risk. Therefore, the most recent date and method by which an individual patient's genotypic record has been updated must be included in the record to ensure that the most timely genotypic risk and population frequency data have been incorporated into the record and to ensure that outdated genotypic records are updated (this assumes an application automatically updates the record, and utilizes a time/date stamp to manage updates) (Figure 2). This notion of state is easily handled by the database in the information system.

Clearly, the management of a genotypic risk database becomes useful, as the central source for determining SNP-specific risk will be managed separately and must be subject to scientific and regulatory oversight. This genotypic risk database will include all known SNPs and their known frequency in the human genome within the population, along with all known health-risk information associated with each SNP.

Genotypic testing considerations

The deployment of genotyping technology in the clinic requires that results from laboratory tests (regardless of the genetic assay platform) be effectively managed for the benefit of patients and the general population. Unlike most laboratory tests used in the clinic, the results of genotyping tests must be stored in a patient-specific database (utilize patient identifier) owing to the large number of potential data points (SNPs) from a single test, as well as contribute to a population-scale database (anonymous identifier). Clearly, the first application of genotyping technology is aimed at surveying drug-metabolizing enzymes to identify patients who are deficient in drug-metabolizing activity, which leverages knowledge that specific SNPs are known to confer this phenotype and testing will be limited to these SNPs. The overarching logic to this approach is that a specific SNP is first associated with a clinically relevant phenotype, and is then deployed as a clinical test. However, the association of known SNPs with clinically relevant phenotypes can (and must) also be determined retrospectively. The population-scale database will reflect the growth of both the number of patients (people) contributing to the genotype database, and the number of SNPs assayed from each person's genome, and will ultimately represent a resource linking genetics with public health informatics. In this approach, a collection of known SNPs is assayed and stored in a population-scale database, which also includes (anonymous) data from the patient's healthcare record. This provides a resource (database) to discover linkage between specific SNP(s) and clinically relevant phenotypes, ultimately linking genotypic data to specific phenotypes.

The data captured from clinical genotyping must include patient identification, genotypic data and other aspects associated with patient-specific sampling, but also accommodate the integration of genotypic data not collected in earlier genotyping tests, information regarding the testing method, quality-control data, as well as the emergence of new technologies involved in testing and data management. Finally, the data must be integrated with a supporting (dynamic) database system that communicates health risks associated with each genotype. Given that the emergence of disease and drug adversity risk with each genotype may be dependent on other genotypic/phenotypic factors, or may simply not yet be known or fully understood, the conversion of genotypic data to health risk must be separate from the patient genotypic data record. The following is a list of data requirements for the genotypic data record:

- Patient identifier
- Sample source/tissue

- Age of patient at sampling
- Genotypic data
- Genotyping/laboratory method
- Quality-control method
- Ethnicity, gender and existing genetic considerations
- Most recent date (and method) of risk-data integration

In addition to the patient's identifier, data must include the source of the genetic material being tested. Potential genetic factors may be tissue-specific, such as genetic variability associated with oncogenesis (e.g., normal tissue vs cancerous tissue), which are crucial, if not the motive, for genotyping. In addition, contaminating genetic material (e.g., bacterial, contaminating human genetic material) may be present

Figure 2. Example of the SNP-specific risk components of a patient's genotypic data, and how they may change using updates that reflect new discoveries from linkage studies.

A

```
[Patient_ID: John Doe]
[Most Recent Update: 11/6/2008]
[Most Recent Update Method: Genosoft 5.6]
[Most Recent Update Source: NIH Human SnipRisk Database, release 263.2]
{Genotype Data}
[SNP Position: ID_1] [Zyg: A, A] [Popul_Freq: 22%] [Risk: Unknown]
[SNP Position: ID_2] [Zyg: T, G] [Popul_Freq: 09%] [Risk: Unknown]
[SNP Position: ID_3] [Zyg: C, C] [Popul_Freq: 78%] [Risk: Unknown]
[SNP Position: ID_4] [Zyg: T, T] [Popul_Freq: 94%] [Risk: Unknown]
[SNP Position: ID_5] [Zyg: G, G] [Popul_Freq: 48%] [Risk: Unknown]
[SNP Position: ID_6] [Zyg: A, C] [Popul_Freq: 32%] [Risk: Cardio, LOW]
[SNP Position: ID_7] [Zyg: C, T] [Popul_Freq: 70%] [Risk: Unknown]
[SNP Position: ID_8] [Zyg: T, T] [Popul_Freq: 29%] [Risk: Unknown]
.
.
.
```

B

```
[Patient_ID: John Doe]
[Most Recent Update: 11/6/2008]
[Most Recent Update Method: Genosoft 5.6]
[Most Recent Update Source: NIH Human SnipRisk Database, release 263.2]
{Genotype Data}
[SNP Position: ID_1] [Zyg: A, A] [Popul_Freq: 22%] [Risk: Unknown]
[SNP Position: ID_2] [Zyg: T, G] [Popul_Freq: 09%] [Risk: Drug, HIGH]
[SNP Position: ID_3] [Zyg: C, C] [Popul_Freq: 78%] [Risk: Unknown]
[SNP Position: ID_4] [Zyg: T, T] [Popul_Freq: 94%] [Risk: Unknown]
[SNP Position: ID_5] [Zyg: G, G] [Popul_Freq: 48%] [Risk: Unknown]
[SNP Position: ID_6] [Zyg: A, C] [Popul_Freq: 32%] [Risk: Cardio, LOW]
[SNP Position: ID_7] [Zyg: C, T] [Popul_Freq: 70%] [Risk: Unknown]
[SNP Position: ID_8] [Zyg: T, T] [Popul_Freq: 29%] [Risk: Unknown]
.
.
.
```

Note that this figure is an amalgam of different data types to demonstrate the behavior of the system, and is not meant to represent how information is presented to the healthcare professional and/or patient. **(A)** John Doe's genotypic component of the electronic health record. **(B)** The electronic health record update using the risk database incorporating a new discovery at SNP position '2'. In this example, SNP position '2' has been linked to an adverse drug response risk.

in skin samples, or mucosal secretions may be considered as a component of the quality-control methods, and can be captured in the sample source data. In addition, the age of the patient is needed for genotypic comparisons made for the patient later in life. As mentioned earlier, many methods for genotyping already exist, and the emergence of new technologies in this arena is certain. Therefore, the method used for a specific data collection/test must be captured, as well as the testing laboratory, personnel involved, and any other relevant information regarding the location and technology employed. The methods employed to ensure that the sample and the laboratory test was performed correctly will contribute to a quality-control determination, and will utilize both genomic sequence and assay standards added to the sample under investigation. Knowledge of an existing genetic condition, such as trisomy 21, will result in triploid data (rather than the expected diploid data) for all genotypic data derived from genetic material on chromosome 21. Finally, given the proposed paradigm that allows the genotypic record to be updated with new risk information, the date of the most recent comparison between the patient's genotypic record and the risk database must be stored (in the patient's record) to ensure risk assessment is based on all data available.

Genotypic data standards & data sources

The data relevant to a patient's genotype will include nucleotide base identification and zygosity at each SNP position, and could include flanking genomic sequence information (depending upon the technology employed). For example, using DNA microarray technology for genotypic screening will essentially be limited to homozygous or heterozygous data for a given SNP position, while genotypic data derived from direct DNA sequencing will provide potentially hundreds of bases of DNA flanking one or more SNPs, which represents a large string of DNA sequence that can be captured. The genotypic data capture must be recognized within the context of the technology or method utilized, and the method or technology utilized must be identified within the genotypic data record (see Figure 3).

This is not meant to infer that any given method is more sensitive or specific, but rather that results are sometimes technology or method dependent. This is somewhat analogous to the utilization of PET and MRI, where results from both tests provide similar insight into the phenotype (phenomena), yet the actual laboratory results are derived from distinct methods. In the case of DNA sequencing, or genotypic data derived from more data-rich sources, the DNA sequence data must be pared down to the SNP(s) that are present (maintained) in the database of risk linkages. Thus, the method of genotyping includes both a categorical description of the biotechnology component (in this case, capillary electrophoresis) and a raw data analysis component (conversion of fluorescent-specific peaks to DNA sequence, and elimination of DNA sequence that does not constitute SNP data). Instances where a given patient harbors a rare genetic condition that is not amenable to SNP-level data must be considered as additional information of the patient, and not a component of a system-wide genotypic data record format.

Genotypic information system

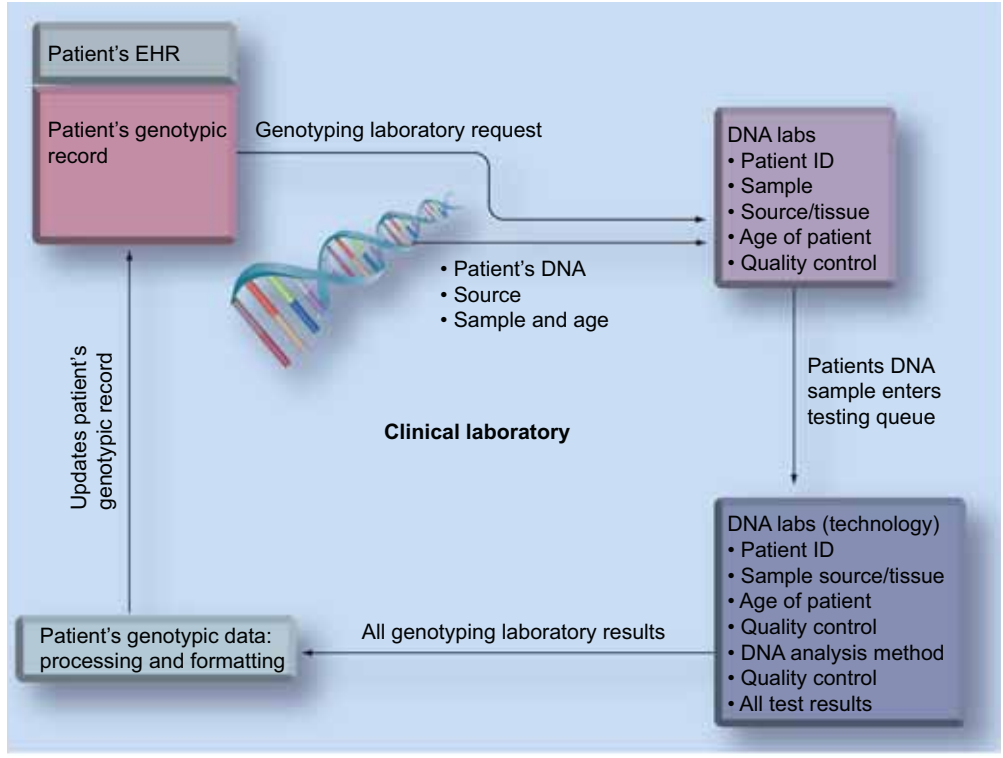
The general architecture of the clinical genotyping information system is represented in Figure 4. The process of DNA testing is described in Figure 3, ultimately deriving or updating patient-specific genotypic data. Once the patient's record has been updated, the data are available for contribution to the HGD. As mentioned earlier, the HGD represents a key source for human genetic research capable of establishing new levels of risk to all known SNPs.

In addition, once the patient's record has been updated the system accesses the risk database to determine if the patient's updated SNP profile includes specific genotypes associated with a known health risk. Some level of overall health risk will be established, which will likely include categorical classifiers such as either 'common' (benign or unknown risk), 'drug' (adverse drug risk) or 'health concern' (some level of overall health risk). These categorical definitions of risk will likely have a simple quantitative component (e.g., low, moderate or high risk) that will be used by the clinical system to flag the attention of healthcare workers and other system components.

Conclusion

Many factors will influence if and how people will derive their genotypic information, including: genotyping test costs, privacy and ethics, as well as the overall cost–benefit of genotyping information. The cost–benefit of genotypic information is dependent upon the rigor of

Figure 3. Data and information flow between the clinic, DNA testing laboratory and secured health-records databases.

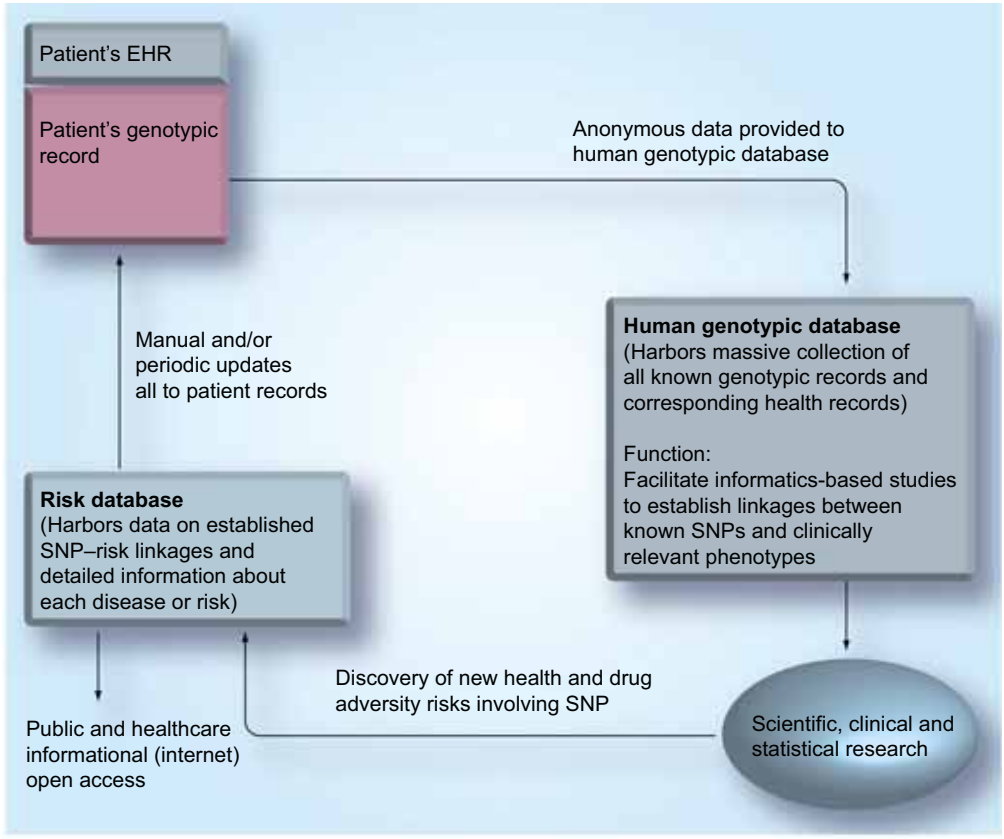


To accommodate the diverse technologies capable of supporting genotyping, data flowing out the clinical DNA laboratory will include data processing, to remove data that is not relevant to the patient's genotypic record, and formatting, to adhere to data storage and security standards.
EHR: Electronic health record.

predicting clinically relevant phenotypic traits based on SNP data. Definitive genetic testing may be tenuous given that every nucleotide in the genome is (theoretically) subject to variance, yet the current strategies for genetic testing are limited to testing for the most common mutations that are known to confer a health risk. For example, there are over 900 mutations in the human genome that have been shown to cause cystic fibrosis (CF), yet most genetic testing laboratories limit their testing to the six most common mutations, and have a predictive success rate of 90% in Caucasians [7]. Using current genetic testing systems, it is not feasible to test for all known mutations that cause CF, given that the benefit of predicting or diagnosing CF from a genetic test does not justify the costs associated with testing hundreds of known mutations from a patient's sample, and that there is a chance that a (rare) specific polymorphism that has not yet been characterized could cause CF and would not be detected in a large-scale genetic screen. It can be expected that any genotyping strategy will be sensitive to false-negative results given that rare SNPs that are not tested under a given genotyping screen may confer a health-risk phenotype.

Deriving sufficient patient information for a large-scale clinical genotyping system will initially involve a large population of patients with mature healthcare records that contain information regarding age-related conditions and diseases, where patient-specific genomic information can be added upon sampling/testing. Ideally, a near-term implementation of clinical genotyping will involve the addition of patient-specific genomic data to an existing healthcare information management system. Certainly, there are many established healthcare groups and systems that are well positioned to benefit from the proposed near-term clinical genomics systems, and partnering with one or more of these groups will both leverage the data and resources inherent to that system and reduce implementation costs by reducing system redundancies. For example, the Veterans Administration

Figure 4. System-wide integration of the patient's electronic health record and its genotypic components, the Human Genotypic Database, research activities, the risk database and the internet.



EHR: Electronic health record.

hospital's healthcare information management system allows for patients to be screened for drug–drug interactions, patient allergies, past medical history and so on. Incorporation of the genomic database into this type of healthcare information-management system would allow pharmacists point-of-care access to genetic information that will be beneficial in making therapeutic decisions. The Veterans Administration system further has a limited drug formulary and a captive patient population that lends itself well to β-testing the clinical genomic system. By

Executive summary

- Hindrances to a societal-scale implementation of clinical genomics include: cost management and established data standards in DNA-analysis technologies, security and access controls in information management, physician and pharmacist education, point-of-care utilization of genomic information, capitalism and return-on-investment issues within DNA screening, and the establishment of valid linkages between allelic markers and healthcare outcomes.
- Overcoming the ethical concerns of healthcare consumers can involve patient-controlled access to categorically defined allelic variants such as 'adverse drug response risk' (i.e., guidance on drug and dose selection) versus 'general health risk' (i.e., risk of specific disease development).
- Near-term implementation of personalized medicine will include the contribution of anonymous records, both genetic and phenotypic, to a database for establishing linkages in humans. The results of these data-intensive association studies will contribute to a 'risk' database that can be accessed in the clinic to translate patient DNA screening results to identifying health outcome concerns and provide guidance on therapeutics.
- Deriving patient DNA samples must include extended metadata from patient-specific information, biological sampling methods and DNA testing technology methods.

starting with a small population, we can then move to the large-scale clinical genomic system to be implemented not only in hospitals but also in other pharmacy practice settings. In addition, this scale-up approach supports a systematic method for assessment of the economic benefits of pharmacogenetic screening [8,9], which is paramount to adoption from private third-party insurance providers (health maintenance organization, preferred provider organizations) as well as Medicare/Medicaid. In conclusion, the implementation of a drug safety program that utilizes genomic data to improve patient care and safety, while at the same time facilitating the movement of clinical genotyping from bench to bedside, will improve general healthcare outcomes.

Financial & competing interests disclosure

Aspects of this perspective article were submitted to the Department of Health and Human Services in response to an RFI report in 2007. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Bibliography

1. Vizirianakis IS: Clinical translation of genotyping and haplotyping data: implementation of *in vivo* pharmacology experience leading drug prescription to pharmacotyping. *Clin. Pharmacokinet.* 46, 807–824 (2007).
2. Swen JJ, Huizinga TW, Gelderblom H *et al.*: Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med.* 4(8), E209 (2007).
3. Gurwitz D, Lunshof JE, Dedoussis G *et al.*: Pharmacogenomics education: International Society of Pharmacogenomics recommendations for medical, pharmaceutical, and health schools deans of education. *Pharmacogenomics J.* 5, 221–225 (2005).
4. Meur Y, Djebli N, Szelag J *et al.*: *CYP3A5*3* influences sirolius oral clearance in *de novo* and stable renal transplant recipients. *Clin. Pharmacol. Ther.* 80, 51–60 (2006).
5. Vizirianakis IS: Improving pharmacotherapy outcomes by pharmacogenomics: from expectation to reality? *Pharmacogenomics* 6, 701–711 (2005).
6. Thornton M, Gladwin A, Payne R *et al.*: Automation and validation of DNA-banking systems. *Drug Discov. Today* 10, 1369–1375 (2005).
7. Gregg AR, Simpson JL: Genetic screening for cystic fibrosis. *Obstet. Gynecol. Clin. North Am.* 29(2), 329–340 (2002).
8. Payne K: Towards an economic evidence base for pharmacogenetics: consideration of outcomes is key. *Pharmacogenomics* 9, 1–4 (2008).
9. Shabo Shvo A: How can the emerging patient-centric health records lower costs in pharmacogenomics? *Pharmacogenomics* 8, 507–511 (2007).
10. Nakajima M, Yokoi T, Mizutani M, Kinoshita M, Funayama M, Kamataki T: Genetic polymorphism in the 5´-flanking region of human *CYP1A2* gene: effect on the CYP1A2 inducibility in humans. *J. Biochem. (Tokyo)* 125(4), 803–808 (1999).
11. King BP, Khan TI, Aithal GP, Kamali F, Daly AK: Upstream and coding region *CYP2C9* polymorphisms: correlation with warfarin dose and metabolism. *Pharmacogenetics* 14(12), 813–822 (2004).
12. Dai D, Tang J, Rose R *et al.*: Identification of variants of CYP3A4 and characterization of their abilities to metabolize testosterone and chlorpyrifos. *J. Pharmacol. Exp. Ther.* 299(3), 825–831 (2001).

Website

101. Genomics at FDA: Regulatory Information www.fda.gov/cder/genomics/regulatory.htm