The Integration of Phenotypic and Genotypic Data:

A Systematic Review and Evaluation of Interoperability Models

Capstone Project

By

Ali A. Al Sanousi, MD

Biostatistics, Epidemiology & Scientific Computing Department

King Faisal Specialist Hospital & Research Centre

Presented to the Department of Medical Informatics, Clinical Epidemiology and Outcomes Research , College of Medicine, Oregon Health and Science University in partial fulfillment of the requirements for the degree of Master of BioMedical Informatics

Spring 2007

Advisor

Christopher Dubay, PhD

School of Medicine

Oregon Health & Science University

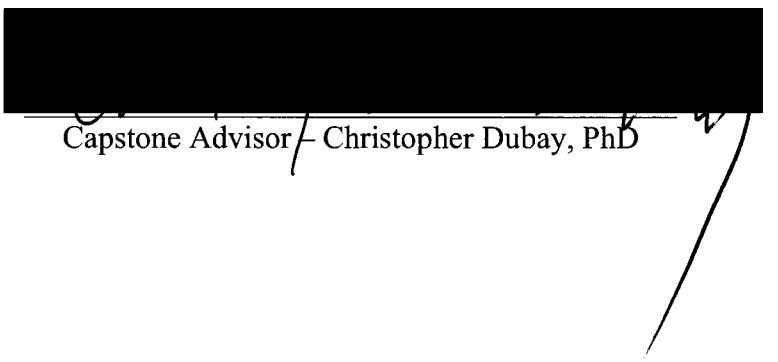
Certificate of Approval

This is to certify that the Master's Capstone Project of

Ali A. Al Sanousi, MD

"The Integration of Phenotypic and Genotypic Data:
A Systematic Review and Evaluation of Interoperability Models"

Has been approved



Capstone Advisor – Christopher Dubay, PhD

PREFACE

The integration of clinical data generated from a patient's bedside, with related genotypic and derived functional genomics data generated from the bench (laboratory), allows presentation of a unified platform for healthcare. The outcome from this integration of data in the next generation of electronic biomedical records will be the harvesting of health information from a true knowledge-based personalized medical record, which is interoperable with other health care information systems.

There are numerous initiatives worldwide to link phenotypic (e.g. clinical) and genotypic data in this way into a contextual model. Currently these projects mainly deal with the solutions to integrate such data for research proposes, and hopefully in the future, they will link to health care information systems, such as Electronic Medical Records (EMRs), to support improved clinical decisions. Until now, there has been no study in the literature that systemically reviewed these models. This capstone presents a method for systematic review of these models to generate a formal and valid set of results which can be utilized for future research. The method will be presented as a proof-of-principle on a set of models. It is planned that this initial work will be extended to a manuscript for publication.

I have chosen this topic to strengthen my knowledge by focusing on this rapidly growing area of scientific research, and to expand this work towards a PhD dissertation. The Biomedical Informatics Network (BioMedMatrix: http://www.biomedmatrix.org) will be the ultimate source for the project findings and the networking tool for researchers who are interested in this field. If you have any questions or you have an interest in the topic, please contact Ali Al Sanousi at (Email: sanousi@kfshrc.edu.sa) for networking.

DEDICATION

This work is dedicated to my wife. Without her tireless Encouragement, I would have given up long ago.

and

To my institution KFSH&RC
which has given me the opportunity to explore the science of biomedical informatics.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my capstone research project supervisor Dr. Christopher Dubay, an Assistant Professor specializing in Bioinformatics at Oregon Health & Science University (OHSU) in the Biomedical Informatics Program for his support, guidance, and continuous encouragement. For this, I will always be grateful. I also would like, to express my sincere gratitude to my academic supervisor, Dr. William Hersh, the Director of the OHSU Biomedical Informatics Program for his support, guidance, and continuous encouragement. For this, I will always be grateful. I'm also grateful for Dr. Holly Jimison, for her advice throughout my academic journey. Likewise, I would like to extend my gratitude to Dr. Paul Gorman and Dr. Shannon McWeeney for supporting my research idea. Finally, I will not forget to express my special and deep thanks to the rest of the academic professors, at the program for supporting my first steps in applied biomedical informatics research.

Last but not least, is my deep appreciation to Dr. Sultan Al Sedairy, the executive director of the Research Center at King Faisal Specialist Hospital & Research Centre Institution for supporting the first and the most important steps of this postgraduate academic journey.

ABSTRACT

Background: Bridging the gap between genotype and phenotype is one of the foremost goals of biomedicine, and consequently biomedical informatics. Effective networking between biologists and clinicians to augment biomedical technology with genotypic and phenotypic knowledge will accelerate the translational research process and enhance the possibility of bridging the gap toward an integrative genotype-phenotype system that could be incorporated in patients' personal medical records. **Methods:** The methods for this project included a systematic review of articles describing systems with an integration of phenotypic and genotypic data. A subset of these systems were scored on quality measures. The scoring method assigned points based on a scale of (0, 1-3), in accordance to a modified Matrix Criteria System for evaluation and scaling. SPSS 12.0 and Microsoft Excel were used to carry out the data analysis. **Results:** Based on the literature review, there are 107 different projects worldwide which are related directly or indirectly with the integration of phenotypic and genotypic data. The majority of the projects are centered in USA 37% followed by the UK 14% then by France 7%. Collaborative projects in Europe represented 10% of the worldwide projects. The international community collaborative work constituted 3% of the projects. A detailed analysis was conducted on two of the high scoring projects (PharmGKB and PhenomicDB). Database functions as a major scoring criterion was employed to critically compare these two projects. The database functions category includes user interface, database design, reporting, technology & platforms, and performance. **Conclusion:** Future genomic medicine clinicians will need to use advanced knowledge and well-calculated diagnostic tests in order to provide targeted treatments to their patients on the road to personalized medical care.

Keywords: phenotype, genotype, integration, interoperability, evaluation, systematic review

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF SNAPSHOTS

I. INTRODUCTION

In the forthcoming decades, biomedical information generated by the field of genomics will have major benefits for the prevention, diagnosis and management of many diseases which have been difficult or impossible to control (1). Medical Care will change more in the next 10 years than in the past 1000 years as a result of biomedical information technology. Rapid developments in electronic technology, the Internet, wireless computers, together with all forms of emerging biomedical informatics technology, like clinical genomics, nanotechnology and biomedical artificial intelligence, will radically alter the way health care organizations do biomedical research in the next decade. These driving forces of change will put new demands on health care leaders and researchers.

BACKGROUND

The Value of BioMedical Informatics

BioMedical Informatics (BMI) is defined as "the science underlying the acquisition, maintenance, retrieval, and application of biomedical knowledge and information to improve patient care, medical education, and health sciences research"[1]. The core component disciplines of BMI include several subspecialties, which are bioinformatics, medical informatics, clinical informatics, imaging informatics and public health informatics (Figure. 1). The spectrum of BMI science intersects with similar in-scope fields including biology, medicine, mathematics, physics, biochemistry, biostatistics, epidemiology, computer science, and management information systems. Therefore, the Clinical Informatics expert can easily apply the scope of BMI, which involves applied informatics research of molecular cellular processes (Bioinformatics), to

[1] http://faculty.washington.edu/gennari/MedicalInformaticsDef.html

tissues and organs (Imaging Informatics), to individuals and patients (Clinical Informatics), to professionals and consumer health education (Consumer Health Informatics), and eventually to the impact on Population and Society (Public Health Informatics).

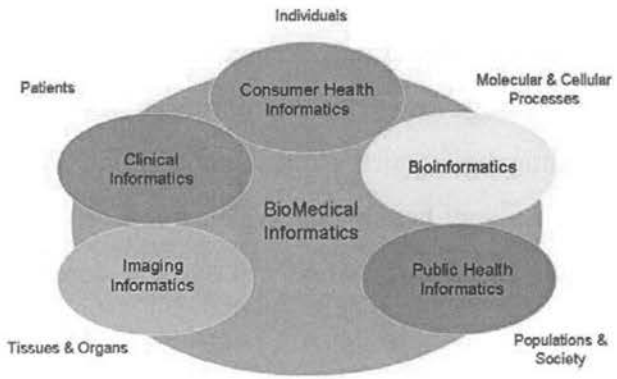


Figure 1. Core Components of BioMedical Informatics.
The spectrum and focus of each component and the impact of each group.

Clinical informatics is one of the most dynamic disciplines of BioMedical Informatics, and it is considered to be a fundamental domain in any successful biomedical informatics program. Advanced Electronic Medical Record systems (EMRs) nowadays include representations of monitoring data, images and DNA samples of patients under surveillance. Thus, investing in biomedical informatics networks is essential by all able countries. It provides the needed infrastructure of a virtual health care data bank, a national health care knowledge base, and a personal clinical health record which are derived from the planning of these strategies (2).

The two domains of biology and medicine have recently merged with the applications of computer science and informatics, into one hybrid discipline: biomedical informatics. There is potential that this merger will yield a wealth of data and information that will enhance our biomedical knowledge toward a better understanding of the

complexity of diseases on the molecular level, and of a better approach to research for disease prevention and treatment through more focused, targeted drugs with higher efficacy and lower toxicity. More biomedical informatics research efforts are needed to build systems that help us in delivering the best possible medical care to our patients and for the early prevention of diseases, resulting in healthy individuals becoming healthier and more productive with a better quality of life. A graphical representation of the intersection between clinical informatics and bioinformatics is illustrated in (Figure 2).

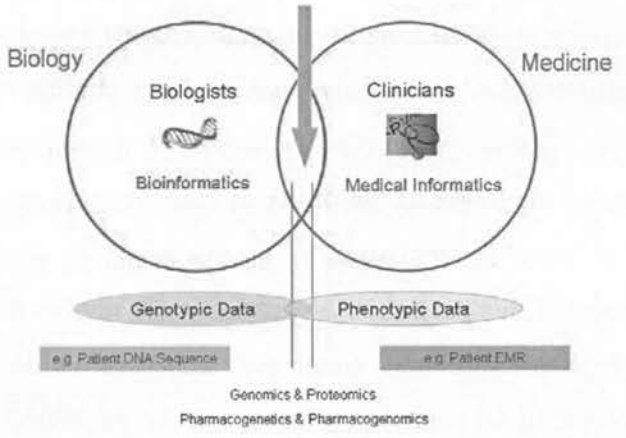


Figure 2. Domains of Biology and Medicine.
The intersection between clinical informatics and bioinformatics, in relation to the practice of clinicians and bio-scientists.

Researchers think that this will lead to new insights in genomic medicine and contribute toward the more efficient and effective use of genomic data to advance clinical care (3). There have been significant efforts to generate funding to stimulate scientists to discover the synergy between medical informatics and bioinformatics, and to propose the best informatics methods to facilitate the delivery of genomic medicine for future health care (4, 5). Some countries are still reluctant to fund research in biomedical informatics, either due to lack of knowledge about its importance or due to lack of cooperation

between biologists and clinicians that may lead to clashes of clinical and basic sciences cultures (6).

Project Scope

Linking genotypic data, including genetic information (e.g. traits, pedigrees) and genomic information (e.g. SNPs, and other characterized, alleles sets, annotations, gene expressions data, etc.) and results of subsequent bioinformatics analyses of that information with phenotypic data, including medical records information (signs, symptoms, investigations results, family history) into an integrated clinical genomics knowledge-based system is in high demand and truly needed. Initiatives of Integration of clinical information at Mayo Clinic[2] and at other similar institutions worldwide have demonstrated promising examples where leading medical and research centers are embracing clinical genomics as the future of the personalized medicine.

The scope of this project is to systematically review the projects that deal with the integration of phenotypic and genotypic data worldwide. Based on the review, a comprehensive list of all available projects will be generated. Finally, evaluative criteria will be introduced to assess these projects.

[2] http://www.mayoclinic.org/state-of-state/future.html

THE RESEARCH QUESTION

The research question which this project aims to answer is how can an integrated pheno-genotypic interoperable system improve medical care, and ultimately the health of our patients?

Hypothesis

There are many international projects that approached the issue of the linking phenotypic and genotypic data into one integral system. These projects remained scattered worldwide and are not currently networked. Many projects have not been recognized. However, we still can learn a great deal from their initiatives and creative work. Creating a network among these projects can further enrich the scientific community and translate into more advanced projects.

PURPOSE OF THE STUDY

The ultimate aim of this project is to integrate clinical data generated by medical records with the related data generated in labs in the area of functional genomics into a unified platform, which is integral and interoperable.

Objectives

The objectives of this project are:

1. To systematically review the literature for models that integrate phenotypic and genotypic data worldwide.
2. To create criteria for evaluation and critically evaluate the obtained interoperability models.

Significance

There are two perspectives to the significance; research and clinical. From the research perspective this work will lead to more in-depth understanding of relationship and further research exploration. And from the clinical perspective this work will facilitate methods of prevention and treatment of diseases, e.g. cancers or complex genetic diseases.

Justification

This research is highly needed in the areas of molecular, genomic and personalized medicine. Likewise, it is needed in the gene therapy research projects, which deal with understanding of the pharmacogenomics associations.

Framework for Evaluation

Since this project will involve evaluations of the obtained projects/models, the evaluated methods were tested against the frameworks for the evaluation information systems. The idea of this research project turned to be highly suitable in accordance to the following frameworks which are adopted from Gorman's Evaluation of Electronic Health Record Systems, chapter 18 of Lehmann's book. Based on the (Chelimsky) classification, the purpose of this project is to systematically review and evaluate the available interoperability models/systems to give insight to professionals; therefore it is classified as a "knowledge evaluation perspective". However, if the study will involve evaluation of the preliminary perceived value of these systems, this may be classified as an "accountability evaluation" as well. And based on (Gremy and DeGoulet) classification the project will highlight on these issues comprehensively. The main perspective of this

project is to guide biomedical informatics professionals toward, linking phenotypic and genotypic information in a unifying standard which embodies a model/system.

In accordance to the Stead's framework, this project will use a fairly appropriate method for evaluation (criteria for analysis), and will be followed by a sequenced and systematic evaluation approach. The NASA (Mankin) approach, classifies our project in the initially stage of feasibility, followed by stage of technology development, and demonstration. In the future, it will also qualify the stage of system development. In addition, based on (Littenberg and Eddy), this project will satisfy the analogous approach in asking questions that must be answered about the underlying system, process, or behavior to be addressed above mentioned levels of systematic evaluation. Lastly, based on the (Oxford/Guyatt), his project will cover, in a structured way, the review and evaluation of the interoperability models. Therefore, it can be classified as a systematic review of models/systems (7).

LITERATURE REVIEW

Information Technology (IT) strategic planning is the key success of any national healthcare system. IT has the potential to improve the quality, safety and efficiency of clinical care. As we approach the genomics medicine era, health care providers plan to increase their investment on IT in the near future because the technology offers the promise of improving the quality of patient care. Many countries are developing national strategies to implement health information infrastructure and Electronic Medical Records (EMRs) into their medical systems. Many of these countries have created research environment to make the future of integrating genetic data into their EMR an aim for medical care advancement.

The advances in clinical information systems to include genetics data on patients, linking the genotypic data with the phenotypic data into a genomic-based EMR facilitating the practice of clinical bioinformatics, will essentially enhance the

quantitative and qualitative methods used in the medical care setting and provide the best available medical care to patients. Researchers of bioinformatics and clinical informatics have incorporated clinical bioinformatics to improve health care, using biological and medical information. An innovative genome-enabled EMR will create opportunities to utilize such integral information in clinical decision, including computerized responses to personalized pharmacogenomics risks (8).

Modern healthcare practice is in need of more objective information on which to base health care decisions, and the accelerating progress and clinical impact of genomics research offers an important source of such information. The convergence of clinical medicine and the life sciences, will result in opportunities in clinical trials and clinically linked medical research (9). Genomics medicine will mandate the use of advance EMR implemented in a research-networking oriented model. Physicians are required to understand the concept of genetic variability, its interactions with the environment, and its implication for patient care. Treating patient through their genetics profiling and prescribing targeted pharmacogenomics medication, will form the shape of our future personalized medicine in the form of Genomics Medicine (10).

BioMedicine Vocabulary

The integration of diverse informatics terminologies is a fundamental prerequisite for the success of any biomedical project supporting personalized medicine. There are a number of structured terminologies used in biomedicine which have been used extensively and have served as a unified coding scheme across informatics systems (11). These include: SNOMED : 1) Systematized Nomenclature of Human Medicine[3]. 2) SNOMED CT: Systematized Nomenclature of Medicine, Clinical Terms. 3) LOINC: Logical Observations, Identifiers, Names and Codes[4]. 4) MIAME: Minimum Information

[3] http://www.snomed.org
[4] http://www.regenstrief.org/medinformatics/loinc

About a Microarray Experiment. 5) UMLS: Unified Medical Language System[5]. 6) GO: Gene Ontology[6]. 7) OMIM: Online Mendelian Inheritance in Man[7], a database that comprise codes for the complex genetic diseases. Online Mendelian Inheritance in Animals[8] (OMIA) also been recently introduced to biomedicine and integrated into Entrez search interface at NCBI to cover up model animals (12).

Model Organisms

Several animal models have been used in biomedical research. For example, the mouse model used in GXD, the Gene Expression Database otherwise called Jax Lab, is good example where genes are studied for tumor markers and correlating these findings with results in humans. Some other model animals used in biomedical research are primates (e.g. Macaca mulatta), rats (e.g. Rattus norvegicus), dogs (e.g. Canis familiaris), forgs (e.g. Xenopus tropicalis) and fish (e.g. Fundulus heteroclitus).

Terms Definition

Phenotype and Genotype:

A phenotype is the "outward, physical manifestation" of the organism. For example, hair or eye colors, IQ, EMR data, cell or medical image, and disease prognosis are all considered phenotypes, A genotype is the "internally coded, inheritable information" carried by all living organisms[9]. PharmGKB defines genotype data as data regarding genomic variants such as single nucleotide polymorphisms (SNPs), insertions and deletions[10]. For instance, DNA sequences, SNPs, transcriptomes, and proteomes are all examples of genotypic information.

[5] http://www.nlm.nih.gov/research/umls
[6] http://www.geneontology.org
[7] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
[8] http://omia.angis.org.au
[9] http://www.brooklyn.cuny.edu/bc/ahp/BioInfo/GP/Definition.html
[10] http://www.pharmgkb.org/submit/genotypeIntro.jsp

Data Integration:

Data integration is considered to be "The movement of data between two co-existing systems. The interfacing of this data may occur once every hour, once a day, etc[11]."

Systematic Review:

A systematic review is a "type of journal article that reviews the literature related to a specific clinical question, analyzing the data in accordance with formal methods to assure that data are suitably compared and pooled[12]".

Evaluation:

System evaluation is "a periodic evaluation of the system to assess its status in terms of original or current expectations and to chart its future direction[13]".

Interoperability:

The Institute of Electrical and Electronics Engineers (IEEE) defines interoperability as: "the ability of two or more systems or components to exchange information and to use the information that has been exchanged[14]." The National Alliance for Health Information Technology (NAHIT) expands a little on the above definitions: "In healthcare, interoperability is the ability of different information technology systems and software applications to communicate, to exchange data accurately, effectively, and consistently, and to use the information that has been exchanged[15]". There are two types of interoperability, syntactic interoperability and semantic interoperability.

Translational Research

The potency of BioMedical Informatics is in the application of bioinformatics tools to the bench side; delivering clinical tools to the bedside and translating these

[11] http://uis.georgetown.edu/departments/eets/dw/GLOSSARY0816.html
[12] Shortliffe. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Glossary.
[13] http://www.answers.com/topic/system-evaluation
[14] IEEE-USA: http://www.ieeeusa.org/policy/positions/NHIninteroperability.html
[15] NAHIT: http://www.nahit.org/cms/index.php?option=com_content&task=view&id=220&Itemid=48

innovations through medical informatics into practice. Thus, translational research is a unique integration of basic and applied sciences that can create a continuum of medical discoveries which will help in advancing our knowledge to innovatively improve our medical care (Figure.2).

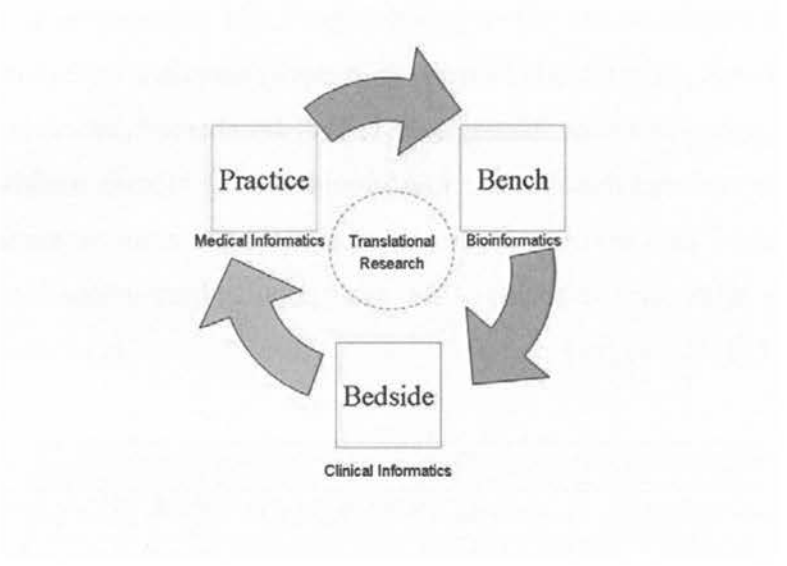


Figure 2. The cycle of Translational research. Bench research to bedside research and translation to practice.

Presently, there has been a fair amount of work done to speed up the feedback form the medical informatics side (innovation in practice) to the bioinformatics side (bench) for further research and discovery. BioMedical Informatics can effectively accelerate the cycle of translation biomedical research. The National Institutes of Health has created an initiative for promoting the concept of translational biomedical research. It is believed that such an initiative will improve the interaction between basic and clinical investigators which will lead in improving the lives of patients afflicted with cancer[16].

A good example of this type of work is the study conducted at the Massachusetts General Hospital related to activating maturations in the EGFR in correlation with the lung cancer to *Gefitinib* therapy, in addition to the study conducted by Dana-Faber Cancer Institute. Both studies managed to merge molecular data with clinical data,

[16] http://ccr.nci.nih.gov/initiatives/TRI

yielding into a translation to clinical practice, which has only taken 90 days to be a standard test used to help patients (13). Thus, the ability to quickly translate discoveries from the bench into clinical orderable test is highly needed.

Biological Databases

Based on a Google Directory[17] search, there are approximately 995 hits when searching the keywords "Biological Medical Databases." Based on the literature review, it is estimated that there are over 850 biomedical databases worldwide. These databases are grouped into division based on their defined focus. A summarized list of these divisions with some example of the most popular international biological databases are provided (Table. 1) (14).

Database Type	Examples (Most popular worldwide)
Ontology	GO, IUBMB, MGED, GMOD, HUGO, HUPO, SOFG
Integral	MSD, EMP, MEROPS, Intger8, GeneCards
Biographical	PubMed, BIOISIS, EMBASE, AGRICOLA, CABI
Taxonomy	NCBI-Toxo, NEWT, ITIS, SP2000, WED
Sequence	EMBL, Genbank, DDBJ, UniPort, Enterz-Protein, RDP
Genes	MGD, FlyBase, RGD, TAIR, MaizeGDB, ZFIN, CGSC
Genomic annotation	Ensembl, Genome Browser, Map Viewer
Clustering	ClusSTr, UniGene, COGs, UniRef, IPI, SYSTERS
Protein classification	CATH, InterPro, PROSITE, PomDom, SMART, PIRSF
Structure	wwPDB, CSC, RESID, NDB, DSSP, HSSP
Expression	ArrayExpress, SMD, CGP, GEO
2D-PAGE	SWISS-2DPAGE, GEO
Interaction	IntAct, BIND, DIP, LIGAND
Enzyme	IntEnz, ENZYME, BRENDA
Pathway	KEGG, BioCyc/EcoCyc

Table 1. International Biological Databases

[17] http://directory.google.com

A novel use of XML is being used in IBM's Genomics Messaging System (GMS) research as part of the Integrated Medical Records (IMR) middleware project. The focus of the GMS design is the representation, transmission, and storage of patient genomic information, particularly in the construction of the unified clinical and genomic record, and exploring the standards required. GMS is a proposed specification for an approach with an emphasis on a specific language for embedding supporting information and management functions in streams of DNA data (20).

According to the Genetic Sequence Data Bank at the National Center for Biotechnology Information (NCBI), it is estimated that there are 56 billion Base Pairs of DNA and 52 Million Sequences in year 2006[21]. This exponential growth represents the outstanding collaboration between biologists that should motivate clinicians as well to collaborate in translational research leading for data integration. Unfortunately, phenotypic data is conceivably the least analyzed form of bioinformatics. Thus, biomedical informatics, as a liaison between the two parties, should promote and support the creation of new technologies to integrate the phenotypic data with the genomic data. This will provide more opportunities for a genome-scale phenotype-genotype correlation (21).

The re-engineering of existing EMRs to include the genomic data integrated in the system has started to happen. This development will lead to a demand on further integration of the health care system towards genomics medicine, optimistically for better quality medical care outcomes. The augmentation of HL7, a health care messaging standard, has also facilitated the introduction of genetic data into the EMR. Health information technology will change the way we look at biomedical research in the very near future.

[21] http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html

In 1909, a Danish botanist Wilhelm Johannsen, coined the word *gene* (using the Greek for "to give birth to"). The terms "genotype" and "phenotype" were created by him in 1911. Genotype refers to the genes an individual has, and phenotype is how those genes are expressed. While genotypic data may determine the presence or absence of a specific disease, phenotypic data is manifested by the gene expression or the manifestations of that specific disease.

According to the "central dogma" of molecular biology as depicted in the figure below, DNA is transcribed into RNA then translated to proteins, which then make small molecules. These proteins then manifested to create the outside characteristics of living creatures (Figure. 3).

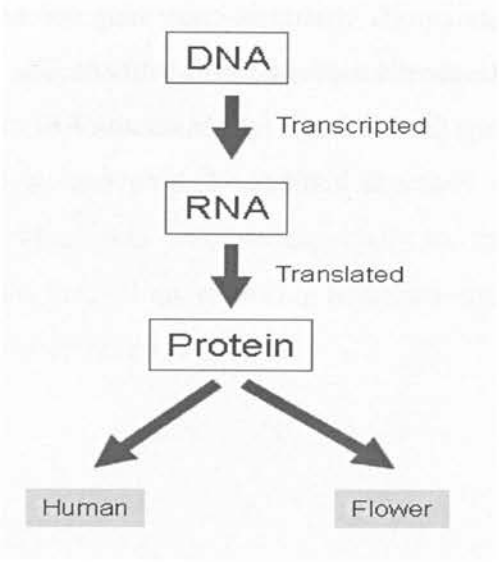


Figure 3. Central Dogma of Molecular Biology

Phenotypic variations occur as a result of the discrepancy in the DNA sequence. Phenotypes are defined differently according to the science it relates to. Genotype can be evidently affected by environmental factors to get altered into a different phenotype, this

phenomenon is called "phenotypic plasticity." The ability of an organism with a given genotype to change its phenotype is an important topic in the genotype-phenotype correlation. Environmental factors also play an important role in the development of a specific genotype to phenotype expression: Genotype → (Development) → Phenotype.

In biomedical sciences, a phenotype is defined simply as the outside physical manifestations which result from the gene inherited from the parents. Linking both phenotype and genotype in biomedical research is a key issue for this project. Phenotypic data are represented in various ways; either in simple textual format as in OMIM, or in a complex quantitative format (text and values). It also can be combinations of both formats. Thus, understanding the nature of complexity of the phenotypic representation is crucial to developing the database system requirement.

Recently, in December 2006, the NIH launched the dbGaP [22], a new database of genome wide association studies. Exploration of the association between specific genes and observable traits can be crucial in enhancing our understanding of diseases and for developing novel diagnostic methods and treatment. The dbGaP started initially with data on studies, including the National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS) and the National Institute of Neurological Disorders and Stroke (NINDS) Parkinsonism Study.

[22] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=gap

II. METHODS

In this research project the integration of phenotypic and genotypic data will be discussed and examined. This work has been conducted in the context of a systematic review for a large sample of related projects worldwide. A database representing a sub-set will be populated, and used in the evaluation of the sub-set using matrix of criteria which is balanced by weight factor and judged by scoring system (Appendix B).

RESEARCH PROCEDURES AND STRATEGIES

Because systematic reviews need a systematic searches, the literature was reviewed on how to conduct a proper systematic review study (22). After that, the research was started by the systematic review approach where the research question was defined and the targeted population was identified. The parameter for comparison and evaluation has also been specified. A literature search was conducted and information sources and searching strategies were defined. Inclusion and exclusion criteria were applied for articles retrieved and project selected. The data were abstracted in a pre-designed Excel template, which was designed especially for this purpose. Data analysis was conducted and methods for polling, exploring heterogeneity, and for assessing biases (i.e., publication biases) was determined (23).

RESEARCH SETTINGS

Area Profile: The intended audience for the results of this project are the biomedical informatics professionals, in particular those whose research focuses on the integration of phenotypic and genotypic data.

Setting and Subjects: Database and Web environment: Systems having individuals with phenotype and/or genotype data.

Study Design: Systematic review and development of criteria for evaluating these interoperability models. Literature review will be conducted as cross-sectional study to all existing project in the literature. Evaluation of the selected websites using the matrix of criteria will be carried out for assessing the quality of these projects.

Inclusion and Exclusion Criteria: All projects worldwide that integrate phenotypic and genotypic data are included in the study. The projects should be interoperable as well. Any project that did not meet this requirement was excluded from the study.

PLAN FOR DATA PROCESSING AND ANALYSIS

Scoring methods will be used to assign points based on a scale of 1-3, in accordance to the modified matrix criteria for Scaling. SPSS 12.0 and Microsoft Excel was be used to carry out the data analysis. A descriptive analysis in form of graphs and tables is presented for different variables. The t-test will be used for continuous variables, and Chi-squared test for the categorical variables. Multivariate analysis with binary outcome variables will be conducted using logistic regression.

DATA COLLECTION AND SAMPLING

Data Collection: The abstraction form used to collect data was adopted from three standardized evaluation forms: 1) Mitretek System Criteria for Evaluation[23]. 2) Ovid Database Evaluation Criteria[24]. 3) River Guide Project Evaluation Matrix[25]. A Check list

[23] http://hitiweb.mitretek.org/docs/policy.html
[24] http://www.ovid.com/site/products/ovidguide/rdidb.htm

will be used to ensure the validity and reliability of the data collected. All forms will be filled by the researcher himself, and a sample will be re-checked to ensure data quality.

Sample: There are 107 pheno-genotypic related projects worldwide. The sample size for this study will be 5% [5.35] of the related projects. A Stratified Random Sampling will be employed. This sampling scheme ensures that our sample is representative of all projects. Strata will be defined by the geographical locations and other relevant factors. Thus, a total of 5 projects will be the sample of this study.

INSTRUMENTS AND DATA ANALYSIS

Criteria for Evaluation: The strategic evaluation criteria used in this project to evaluate the reviewed projects and to assess the quality of different models, was set to be based on the following (Table. 2).

Criteria for Evaluation	
I. Quality of Information	**III. Search Capabilities**
1. Credibility	1. General Search
2. Content	2. Advanced Search
3. Disclosure	**IV. Services and Support**
4. Links	1. Services
II. Database Functions	2. Help
1. User Interface	3. Support
2. Database Design	
3. Reporting	
4. Technology & Platforms	
5. Performance	

Table 2. List of Evaluation Categories and Major Criteria

[25] Willow Brook Partners by Don Fornes

Weighting and Scoring System: The evaluation criteria were weighted and scored by a modified system adopted from the *Axia Consulting* evaluation scoring system[26]. The required weightings, the evaluative scores and the calculated points assigned were set to be as follows:

Requirement weighting factors: Essential (4 x), Important (3 x), Desirable (2 x), or Nice to have (1 x). The evaluation score are based upon reviewing and analyzing the projects: 0 = does not meet requirements, 1 = partially meets requirements, 2 = meets requirements, or 3 = exceeds requirements. This Demy table was deployed on Excel to score data and calculate the points scored (Table. 3).

		Project A		Project B		Project C		Project D	
Criteria	Weight	Score	Points	Score	Points	Score	Points	Score	Points
Criteria # 1	4x	3	12						
Criteria # 2	4x	3	12						
Criteria # 3	4x	3	12						
------	---	---	---						

Table 3. Comparison and Scoring Matrix

The maximum points a project can get is 12. This can only occur if the criterion was weighted as essential (4 x) and the evaluative score revealed to be 3. Project A is an example of how the points are calculated. The points are calculated by multiplying the evaluation score by the weighting factor. The weighting factors were assigned by the researcher based on a literature review and his own judgment as to the importance of a certain criteria. Likewise, the scores also were assigned by the researcher based on in-depth assessment of each project under evaluation. The projects were critically evaluated based on four main criteria: quality of information, database functionality, search capabilities, and service and support.

[26] http://www.axia-consulting.co.uk/html/rfp_evaluation.html

Each major criterion had sub-criteria that had been subject to the scoring system. Scores were assigned to each sub-criteria carefully and fairly. The average of each group of sub-criteria and the cumulative summation of all averages were calculated. Comparatively, the other criteria and sub-criteria were scored.

$$\text{Points} = (\text{weighting}) * (\text{evaluation score})$$

Equation 1. Points Calculation

The average of each sub-criterion then calculated and lastly the average of all criterion of each project under evaluation is also calculated. The project in which its database model achieves the highest points will be among the best models in accordance to the criteria of evaluation.

Evaluation of the Interoperability Models: to evaluate the interoperability models, five projects were selected to deploy the criteria for an evaluation model which was designed as proof-of-concept for this research. The following table contains the project details listed alphabetically (Table. 4).

Name	Country	Description
CGVdb[27]	Taiwan	Chinese Gene Variation Database
HKIS[28](BioPack)	France	Integrated biological and anatomo-clinical system
IGS[29]	UK	Integrated Genotyping System
PharmGKB[30]	USA	Clinical pharmacogenomics knowledgebase
PhenomicDB[31]	Germany	Multi-species genotype-phenotype database

Table 4. List of Projects and Description

[27] http://www.cgvdb.org.tw
[28] http://isoft.free.fr/hkis
[29] http://bioinformatics.well.ox.ac.uk/project-igs.shtml
[30] http://www.pharmgkb.org
[31] http://www.phenomicdb.de

Among the numerous projects, the following were chosen for a variety of the following reasons. First, each represents a different country of origin. Second, they had been published in the literature. Finally, all five projects at this point in time have online databases which are accessible and functional.

Pilot Study: A single tentative evaluation form was filled by random from the project to fix the coding and validities the data analysis on the SPSS spreadsheet. Dummy tables were produced and in accordance to research needed. The forms and data were tested and quality was ensured.

ETHICAL CONSIDERATIONS

The research study did not involve any active or life substance. The databases and downloaded software, which were used, are either open source or trial versions. Proper attributions to the projects and related databases were also included.

LIMITATIONS OF THE STUDY

One of the difficulties with this study is the complexity of the evaluation criteria, especially with the some sub-criterions and the difficulty of applying these to non-standardized project models under study. Another factor adding to the complexity is the diversity of the projects.

III. MANAGEMENT PLAN

This project was conducted to be presented to the Department of Medical Informatics & Clinical Epidemiology at Oregon Health & Science University in partial fulfillment of the requirements for the degree of master of biomedical informatics.

PROJECT ORGANIZATION

This work is supported by the King Faisal Specialist Hospital and Research Center headquarters through a scholarship fund to the researcher. The project findings are to published in biomedical informatics journals for networking and for the use of academic and educational proposes. The final results and data analysis updates will also be posted online on the *BioMedMatrix* website.

Work Plan

The evaluation forms will be filled by the Principal Investigator himself to ensure accuracy. All data will be entered to the computer in the SPSS spread sheet, and analyzed by the researcher. The capstone summary and main results (Tables & Graphs) are expected to be online by the end of 2007.

Research Budget

This work was not funded by any profit or non-profit organization. A domain name was reserved and hosted for data demonstration. Hosting of the project website is a personal proactive effort.

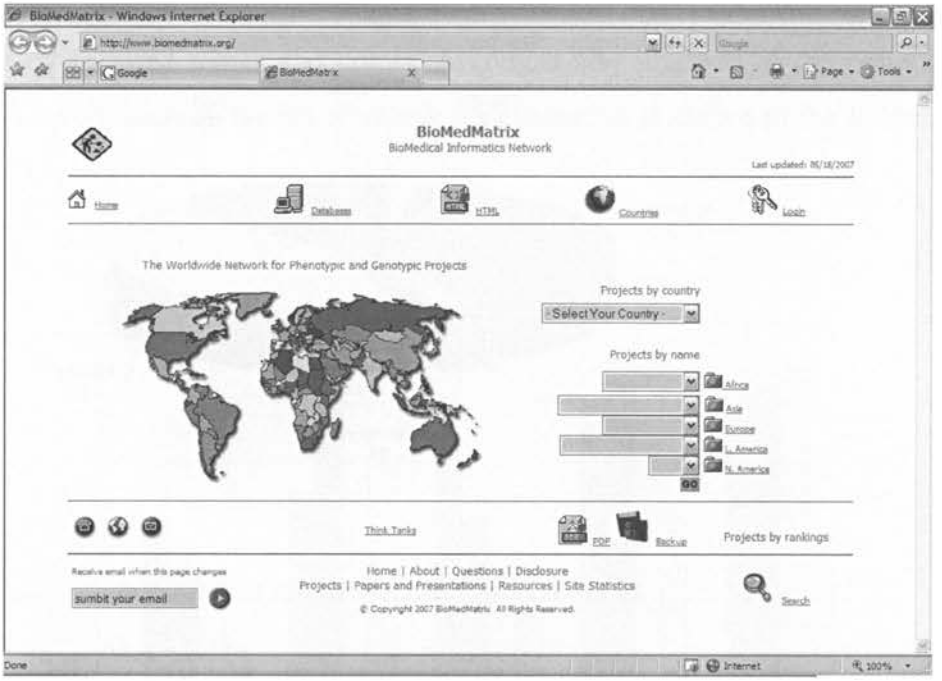
The capstone work has extended though year 2007. The project started in June of 2006 and continued until May 2007.

Activity	1st Phase Preparatory Phase			2nd Phase Pilot Study Phase			3rd Phase Main Study Phase			4th Phase Reporting Phase		
	2006			2006			2006	2007		2007		
	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
- Topic Selection - Literature Review - Advisor Approval	X	X	X									
- Proposal Writing - Data Assessment - Dummy Table Preparation				X	X	X						
- Sampling - Data Collection - Data Entry - Data Analysis							X	X	X			
- Discussions - Writing - Project Daft								X	X	X		
- Submission of the Progress Report											X	
- Submission of the Final Report												X

The plan for dissemination and utilization of results and future updates will be available online at *BioMedMatrix* website (http://www.biomedmatrix.org). A final report of the capstone project will be online for the use of the future researchers who are interested in the field of biomedical informatics. The official website of the project will be developed and maintained by the principle investigator.

BioMedMatrix

The Worldwide Network for Phenotypic and Genotypic Projects



Project website prototype online

www.biomedmatrix.org

Based on the literature review, there are 107 different projects worldwide which are related directly or indirectly with the integration of phenotypic and genotypic data. The five projects which have been selected as a sample for this study are CGVdb, HKIS (AMADEA BioPack), IGS, PharmGKB, and PhenomicDB (Appendix A).

The data analysis has reveled that most of the projects are located in Europe (42%) and North America (39%). The third contender is the continent of Asia, which contributed 11% of the projects. Notably, 3% of the projects are international collaborations (Figure. 5).

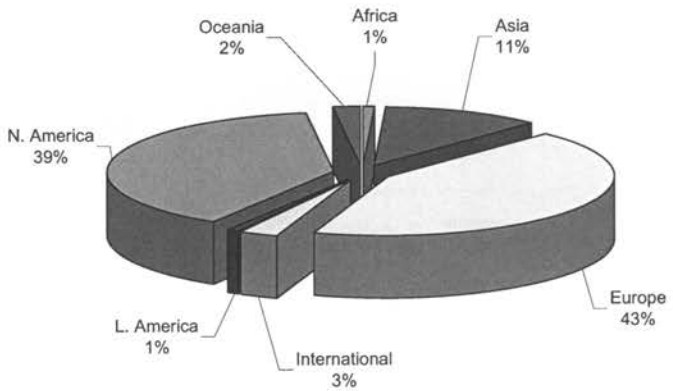


Figure 4. Distribution of Projects by Continent

The majority of the projects are centered in USA 37% followed by the UK 14% then by France 7%. Collaborative projects in Europe represented 10% of the worldwide projects. The international community collaborative work constituted 3% of the projects (Figure. 6).

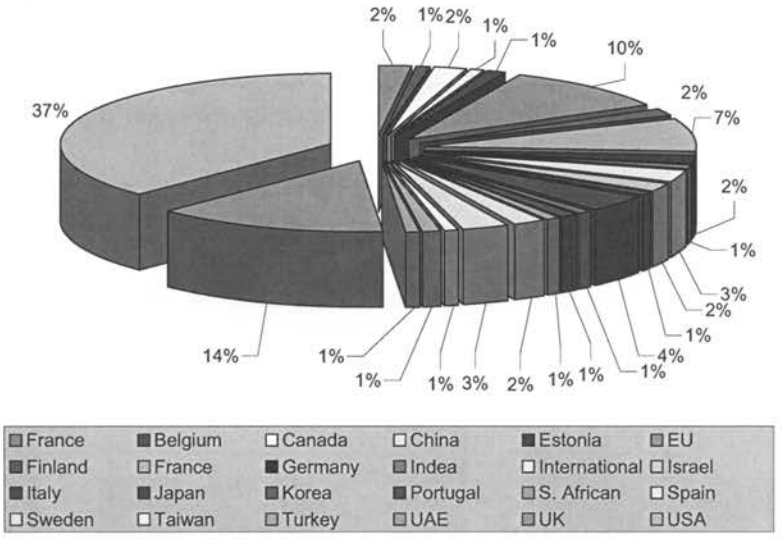


Figure 5. Distribution of Projects by Country

These projects were assessed against a set of criteria for evaluation which was weighted by factored scores and points. A comparative analysis of the performance of each project representing the main criteria for evaluation is shown in the following chart (Figure. 7).

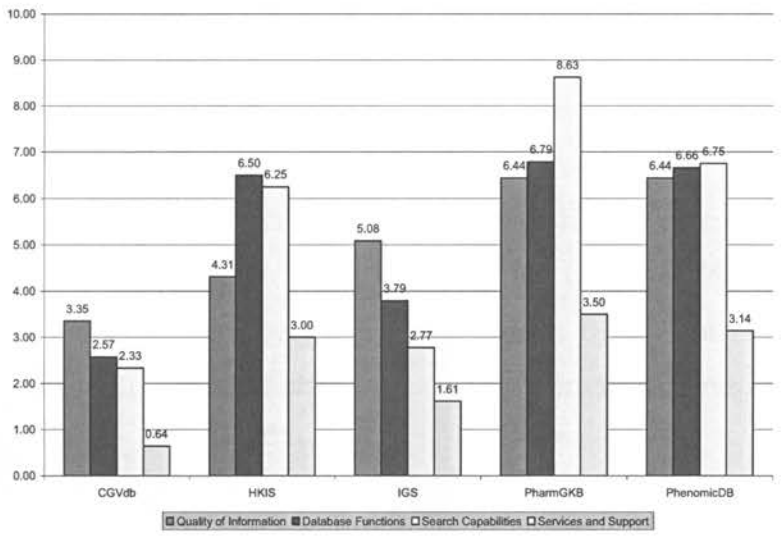


Figure 6. Projects Evaluation by Points Scored

A detailed analysis was conducted on the two projects (PharmGKB and PhenomicDB) which showed a higher performance that the rest of the projects. Database functions as a major category of criteria was employed to critically compare these two

projects. The database functions category includes user interface, database design, reporting, technology & platforms, and performance (Appendix B). Each of these main criteria has sub-criteria premeditated to critically evaluate biomedical databases (Figure. 6).

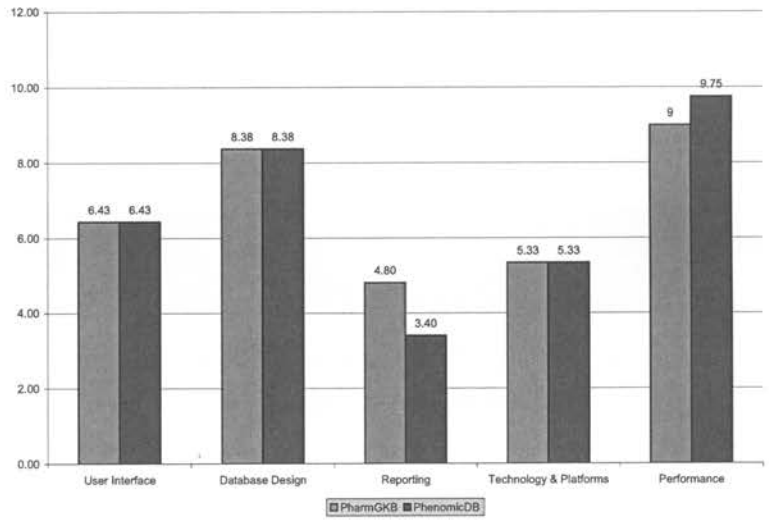


Figure 7. Analysis of Database Functions

On the whole, PharmGKB scored the highest points 87.44 comparing to the rest of the projects with a difference of 5.49 points from the competing PhenomicsDB project (Figure. 8).

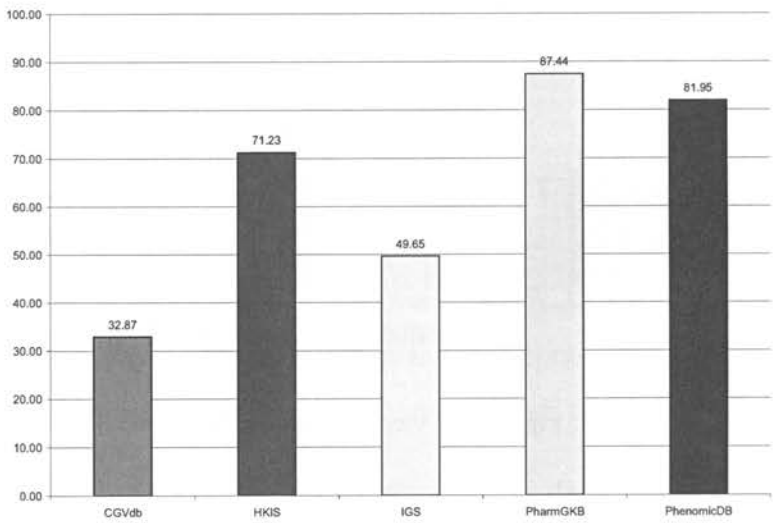


Figure 8. Overall Projects Evaluation

international community will move toward the human health and patient care. The genomic impact on health care industry at the levels of population, disease, patient, tissue and organ banks, cellular biology and down to the level of genes giving specific risks.

Furthermore, when projects assessed using the pre-designed criteria for evaluation they appear to vary in the levels of quality of information, database functions, search capabilities and support they provide. Having said that, projects like HKIS (AMADEA BioPack) and PhenomicDB were nearly equal in the database functions. Similarly, PhenomicDB project has just about the same services and user support in compression to PharmGKB project. On the contrary, the three European project HKIS (AMADEA BioPack), IGS and PhenomicDB have shown significant discrepancy in quality of information, database functions and search capabilities. In addition, notably PharmGKB and PhenomicDB projects got considerably higher scores in all evaluated criteria than others and since the database functions category is considered to be critical criteria for excellence therefore, they were chosen to be analyzed intimately. Whereas both projects revealed almost exactly similar quality in the user interface, database design and in technology and platforms, but they were different in reporting more prominently in the database performance. PharmGKB project evidently scored higher in database performance that PhenomicDB.

Finally, the overall projects evaluation analysis revealed that the PharmGKB project scored the highest points comparing to the rest of all projects under evaluation in all sub-criteria and accordingly all main criteria, distinguishing in database functions and search engines capabilities.

Major Challenges

Moving from "one drug fits all" to personalized therapy is one of the outcomes of the integrated clinical systems that correlate phenotypic and genotypic data. Giving the right treatment at the right dose for the right person at the right time will perceptibly lead to the right outcome. The ultimate goal of personalized medicine is to translate genomic

data into a specific knowledge about the related disease to enable the informatics oriented clinicians to diagnose genetically predisposed conditions, then recommend personal preventive measures and prescribe targeted effective medications which are efficient and least likely to cause adverse reactions the individual (28).

One challenge in creating a phenotypic information system is the fact that the phenotypic clinical presentation of patients is changeable over time. Hence, as humans gets older, their clinical manifestations will also change. There are sets of phenotypic parameters that need to be managed carefully in order to get a robust database that can be linked with the genotypic databases on the other side. It has been a challenge to represent phenotypes to be ready for phenotype-genotype correlation in biomedical research. Despite that, it is the avenue for personalized medicine which the global community is targeting. (29).

VI. CONCLUSION

Future genomic medicine clinicians will need to use advanced knowledge and well-calculated diagnostic tests in order to provide targeted treatments to their patients on the road to personalized medical care. Since prevention is better than cure, preventive molecular medicine will reduce hospitalization and unnecessary invasive procedures, hopefully improving quality of life and reducing medical care expenses (30).

Worldwide, and particularly in the Eastern Mediterranean region were the rate of consanguinity is high, an in-depth understanding of complex genetic diseases will help in resolving much of human suffering (31). The advances in the human genome project will mandate the integration of the patient's genomics data into the EMR to improve the treatment of such complex genetic diseases. There are several ways to approach the genotype-to-phenotype relationship with pharmacogenomics discovery (32). Likewise, there are different ways integrate genomic data in EMRs. However more integration efforts have to be made and the ideal way to best serves patients' health and to advance research has yet to be discovered (33) .

The vision of this project is to assess projects which develop pheno-genotypic databases for biomedical informatics research. Since the Human Genome Project was accomplished, there has been a need for an in depth standardization of vocabulary and terminology to facilitate the integration of phenotypic and genotypic data across information systems. A pheno-genotypic database which is designed to adopt structured knowledge over time is desired. The information system should be user friendly, integral, accessible, and achievable. Information in the system should be retrievable and easily queried (34). Sufficient security measures should be implemented and deployed to make sure that the data in the database is intact and ready to use for biomedical research to serve mankind.

VII. APPENDICES

APPENDIX A

Worldwide Phenotype-Genotype Projects: listed alphabetically.

1. Access Grid
2. AGRE
3. ARTEMIS
4. BioAlma
5. BioGRID (Japan)
6. BIOGRID* (UK)
7. BIOPATTERN
8. BIRN
9. BodyMap-Xs
10. caBIG™ BMI Grid
11. CardioGenomics
12. CARDITIS
13. CFTR
14. CGAP
15. CGVdb
16. CLEF
17. CoreGRID
18. CTGA Database
19. dbGaP
20. DeCode
21. DEG
22. DEISA
23. DEVASPIM
24. DMID
25. DOE Genomics
26. DynaMetrix
27. Ecgene
28. EDCTP
29. EGAPP
30. EGEE
31. EGP
32. ENCODE
33. Ensembl
34. Entrez Gene
35. ESTHER
36. euGenes
37. EuroGentest Project
38. European DataGrid
39. eVOC
40. ExPASy
41. EXPOLDB
42. FatiGO
43. FDD
44. FlyBase
45. GAIN
46. GDB
47. GDPC
48. GDPInfo
49. GEANT
50. GEMSS
51. GENATLAS
52. Gene Cards
53. Gene Ontology
54. GeneLynx
55. GenePool
56. GeneTests
57. GMS-IBM
58. GPDB
59. Grid Operations Centre
60. GRIP
61. GriPhyN
62. GSV
63. GTL
64. HapMap
65. HCT Project
66. Health GRID
67. Helix
68. HGMD
69. HGNC
70. HGVbase
71. HKIS
72. HPRD
73. HuGENet
74. IGS
75. IMG
76. IMGT
77. Immunogrid
78. INFOBIOGEN
79. INFOBIOMED
80. INFOGENMED
81. IPG
82. IRIS
83. IRSA
84. KEGG
85. LINK3D
86. Lussier Research Group
87. MAGNet
88. MAMMOGRI
89. Medgene
90. MEMO
91. MGI
92. MGS
93. myGrid
94. OBO Foundry
95. OMIM
96. Open HER
97. P3G Knowledge Database
98. PharmGKB
99. PhenomicDB
100. PHGU
101. PING
102. PoCT
103. PRIDEH-GEN
104. RGD
105. RZPD
106. SEMANTIC MINING
107. SHGC

APPENDIX B

List of Criteria for Evaluation and Weight Factor Assigned.

Criterion	Weighting Factor (0, 1-4)
I. Quality of Information	
Credibility	
Trusted Source	4
Credentials (Investigators)	3
Absence of Bias	2
Context	2
Currency	2
Editorial Review Process	3
Content	
Accuracy	4
Hierarchy of Evidence	4
Original Source Stated	3
Disclaimer	3
Disclosure	
Purpose of the site	2
Profiling Sponsors	3
Links	
Architecture (ease of navigation)	2
Back Linkages and Descriptions	1
II. Database Functions	
User Interface	
Accessibility	3
Logical organization	2
Internal Search Engine	3
Overall intuitive design	1
Guided workflow and navigation	1
Tool tips and help functions	2
Ability to attach documents	3
Database Design	
Connection to data sources: XML, RDBMS, flat files, etc)	3
Access to external data banks (GenBank, GO, OMIM, EMBL, SwissProt, etc)	4
Traceability of data management	2
Data Analysis Accessibility	3
Provision of results	3
Changing parameters & reusing the process	2
Management of very big datasets	3
Management of Metadata	3
Reporting	
Mechanism for Feedback	1

Custom report creation	1
Intuitive graphs and charts	2
Real-time processing	3
Integration to Microsoft Excel	2
Technology & Platforms	
Hosted with a unique domain name	3
Adopted database (e.g. SQL Server, Oracle, Sybase, etc)	3
Ability to work "off-line"	1
Performance	
Reliability	3
Speed of results delivery	3
Currency of results	4
Availability	3
III. Search Capabilities	
General Search	
Keyword	3
Phrase searching	2
Truncation	3
Advanced Search	
Diseases	4
Pathways	4
Drugs	4
Microarray	3
Genotypes	2
Phenotypes	2
Genotypes & Phenotypes Combined	4
Relevant literature	3
IV. Services and Support	
Services	
Download Options	3
Printable format	1
e-mailing results	2
Help	
Contextual help	1
Tutorial within database	2
Comprehensible error messages	1
User guide	1
Support	
Telephone support	1
email support	2
Online help function	1
Total	

VIII. GLOSSARY

1.	IT	Information Technology
2.	BMI	BioMedical Informatics
3.	DNA	Deoxyribo Nucleic Acid
4.	EMR	Electronic Medical Record
5.	GMS	Genomics Messaging System
6.	HGP	Human Genome Project
7.	HL7	Health Level 7
8.	IMR	Integrated Medical Records
9.	IQ	Intelligent Quotient
10.	KFSH&RC	King Faisal Specialist Hospital & Research Centre
11.	LOINC	Logical Observations, Identifiers, Names and Codes
12.	MIAME	Minimum Information About a Microarray Experiment
13.	NCBI	National Center for Biotechnology Information
14.	NCBI	National Center for Biotechnology Information
15.	OHSU	Oregon Health & Science University
16.	OMIM	Online Mendelian Inheritance in Man
17.	RDBMS	Relational Database Management System
18.	SNOMED	Systematized Nomenclature of Human Medicine
19.	SNOMED CT	Systematized Nomenclature of Medicine, Clinical Terms
20.	SNP	Single Nucleotide Polymorphisms
21.	SPSS	Statistical Package for Social Science
22.	UCSF	University of California, San Francisco
23.	UMLS	Unified Medical Language System
24.	URL	Universal Resource Locator
25.	WHO	World Health Organization
26.	XML	Extensible Marked up language

IX. REFERENCES

Bibliography

1. Brundtland GH. Genomics and world health. Report of the Advisory Committee on Health Research. WHO Library Cataloguing-in-Publication Data 2002.

2. Stead WW, Lorenzi NM. Health informatics: linking investment to value. J Am Med Inform Assoc 1999;6(5):341-8.

3. Knaup P, Ammenwerth E, Brandner R, Brigl B, Fischer G, Garde S, et al. Towards clinical bioinformatics: advancing genomic medicine with informatics methods and tools. Methods Inf Med 2004;43(3):302-7.

4. Martin-Sanchez F, Iakovidis I, Norager S, Maojo V, de Groen P, Van der Lei J, et al. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. J Biomed Inform 2004;37(1):30-42.

5. Maojo V, Iakovidis I, Martin-Sanchez F, Crespo J, Kulikowski C. Medical informatics and bioinformatics: European efforts to facilitate synergy. J Biomed Inform 2001;34(6):423-7.

6. Altman RB. The interactions between clinical informatics and bioinformatics: a case study. J Am Med Inform Assoc 2000;7(5):439-43.

7. Lehmann HP. Aspects of Electronic Health Record Systems. Health informatics. New York: Springer; 2006.

8. Hoffman MA. The genome-enabled electronic medical record. J Biomed Inform 2007;40(1):44-6.

9. Rindfleisch TC, Brutlag DL. Directions for clinical research and genomic research into the next decade: implications for informatics. J Am Med Inform Assoc 1998;5(5):404-11.

10. Shah R, Darne B, Atar D, Abadie E, Adams KF, Zannad F. Pharmacogenomics in cardiovascular clinical trials. Fundam Clin Pharmacol 2004;18(6):705-8.

11. Sarkar IN, Cantor MN, Gelman R, Hartel F, Lussier YA. Linking biomedical language information and knowledge resources: GO and UMLS. Pac Symp Biocomput 2003:439-50.

12. Lenffer J, Nicholas FW, Castle K, Rao A, Gregory S, Poidinger M, et al. OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. Nucleic Acids Res 2006;34(Database issue):D599-601.

13. Couzin J. Pharmacogenomics. Cancer sharpshooters rely on DNA tests for a better aim. Science 2004;305(5688):1222-3.

14. Gorman PN. Evaluation of Electronic Health Record Systems. In: Lehmann HP, editor. Aspects of Electronic Health Record Systems. Health informatics. New York: Springer; 2006. p. 18-0:18:18.

15. McLaren CG, Bruskiewich RM, Portugal AM, Cosico AB. The International Rice Information System. A platform for meta-analysis of rice crop data. Plant Physiol 2005;139(2):637-42.

16. Guo J, Takada A, Tanaka K, Sato J, Suzuki M, Suzuki T, et al. The development of MML (Medical Markup Language) version 3.0 as a medical document exchange format for HL7 messages. J Med Syst 2004;28(6):523-33.

17. Koncar M. Implementing the HL7v3 standard in Croatian primary healthcare domain. Stud Health Technol Inform 2004;105:325-36.

18. Lechleitner G, Pfeiffer KP, Wilhelmy I, Ball M. Cerner Millennium: the Innsbruck experience. Methods Inf Med 2003;42(1):8-15.

19. Pagon RA, Tarczy-Hornoch P, Baskin PK, Edwards JE, Covington ML, Espeseth M, et al. GeneTests-GeneClinics: genetic testing information for a growing audience. Hum Mutat 2002;19(5):501-9.

20. Robson B, Mushlin R. Genomic messaging system and DNA mark-up language for information-based personalized medicine with clinical and proteome research applications. J Proteome Res 2004;3(5):930-48.

21. Amanda C. Integration of Genomic and Phenotypic Data. In: Francisco Azuaje JD, editor. Data Analysis and Visualization in Genomics and Proteomics; 2005. p. 83-97.

22. McGowan J, Sampson M. Systematic reviews need systematic searchers. J Med Libr Assoc 2005;93(1):74-80.

23. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. CMAJ 1988;138(8):697-703.

24. Boulakia SC, Lair S, Stransky N, Graziani S, Radvanyi F, Barillot E, et al. Selecting biomedical data sources according to user preferences. Bioinformatics 2004;20 Suppl 1:I86-I93.

25. Fiddy S, Cattermole D, Xie D, Duan XY, Mott R. An integrated system for genetic analysis. BMC Bioinformatics 2006;7:210.

26. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, et al. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J 2001;1(3):167-70.

27. Kahraman A, Avramov A, Nashev LG, Popov D, Ternes R, Pohlenz HD, et al. PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics. Bioinformatics 2005;21(3):418-20.

28. Sadee W. Pharmacogenomics: harbinger for the era of personalized medicine? Mol Interv 2005;5(3):140-3.

29. Lesko LJ. Personalized medicine: elusive dream or imminent reality? Clin Pharmacol Ther 2007;81(6):807-16.

30. Willard HF, Angrist M, Ginsburg GS. Genomic medicine: genetic variation and its impact on the future of health care. Philos Trans R Soc Lond B Biol Sci 2005;360(1460):1543-50.

31. Acharya T, Rab MA, Singer PA, Daar AS. Harnessing genomics to improve health in the Eastern Mediterranean Region - an executive course in genomics policy. Health Res Policy Syst 2005;3(1):1.

32. Altman RB, Klein TE. Challenges for biomedical informatics and pharmacogenomics. Annu Rev Pharmacol Toxicol 2002;42:113-33.

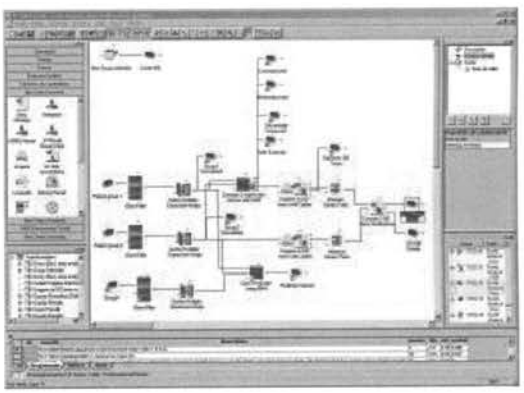
33. Sax U, Schmidt S. Integration of genomic data in Electronic Health Records--opportunities and dilemmas. Methods Inf Med 2005;44(4):546-50.

34. Nadkarni PM. Information retrieval in medicine: overview and applications. J Postgrad Med 2000;46(2):116-22.

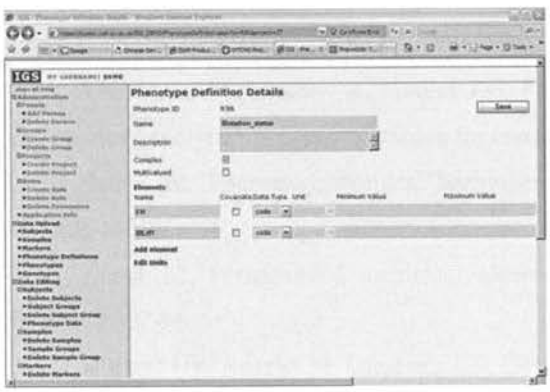
Snapshot 1. CGVdb

Snapshot 2. PhenomicDB

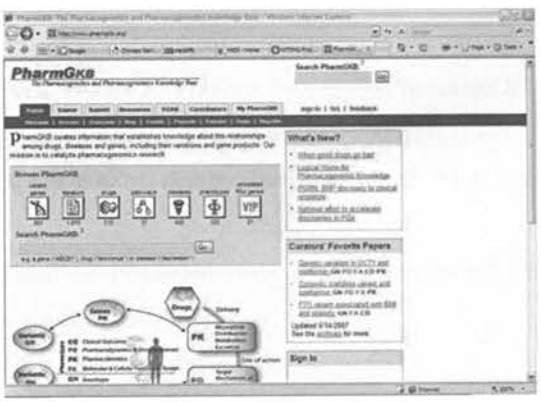
Snapshot 3. HKIS



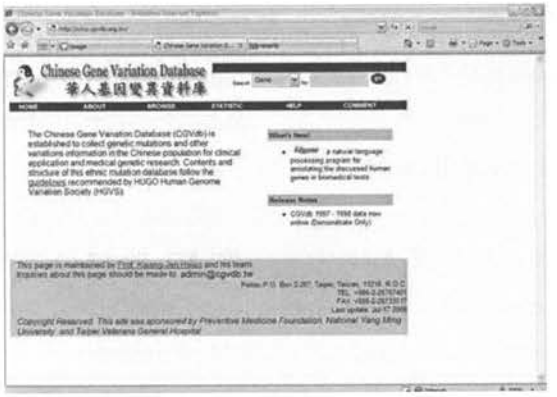
Snapshot 4. HKIS: Amadea BioPack



Snapshot 5. IGS



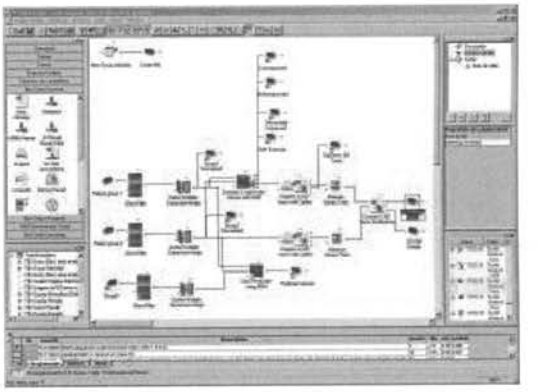
Snapshot 6. CGVdb



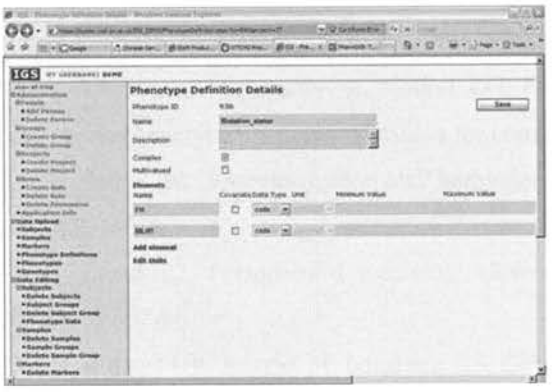
Snapshot 1. CGVdb

Snapshot 2. PhenomicDB

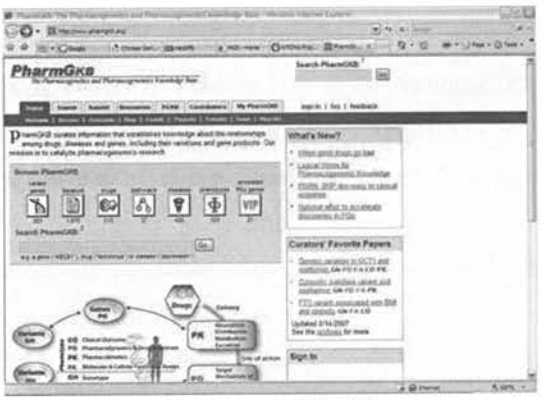
Snapshot 3. HKIS



Snapshot 4. HKIS: Amadea BioPack



Snapshot 5. IGS



Snapshot 6. CGVdb

X. VITA

Ali A. Al Sanousi, MD

Dr. Al Sanousi works as a medical informatics specialist and assistant research scientist at King Faisal Specialist Hospital & Research Centre (KFSH&RC), Riyadh, Saudi Arabia. He is also the program manager of the National Genetic and Birth Defects Registry (NGBDR). After completing his medical degree, MBBS, from the College of Medicine at King Saud University (KSU) in Saudi Arabia, he obtained a diploma in International Health (DIH) from the Royal College of Physicians, Dublin, Ireland. He had training in applied epidemiology and biostatistics from the Center of Disease Control and Prevention (CDC) in Atlanta, Georgia. Dr. Al Sanousi is a licensed Microsoft Certified System Engineer (MCSE) from Microsoft Corporation, and a Certified Internet Web developer (CIW) from the International Web Masters Association. His interest in medical Informatics and clinical genomics, guided him to complete his certificate in medical informatics from Stanford University, Palo Alto, California. Since then, his teaching and research experiences remained focused on health and biomedical informatics, where he developed and conducted several training courses in research methodology for health care professionals. Dr. Al Sanousi also contributed to a non-profit organization as a member of the board of trustees and the chairman of the education committee for the Internet Society, Saudi Arabian Chapter (ISOC-SA). He is currently pursuing his masters' degree in BioMedical Informatics at the Oregon Health and Science University (OHSU), Portland, Oregon. He is a member of several medical informatics professional organizations. His research interest focuses on the biomedical informatics' translational research and the integration of clinical information systems.