

Original Paper

Incorporation of Personal Single Nucleotide Polymorphism (SNP) Data into a National Level Electronic Health Record for Disease Risk Assessment, Part 1: An Overview of Requirements

Timur Beyan, MD, PhD; Ye im Aydın Son, MD, PhD

Informatics Institute, Department of Health Informatics, Middle East Technical University, Ankara, Turkey

Corresponding Author:

Ye im Aydın Son, MD, PhD

Informatics Institute

Department of Health Informatics

Middle East Technical University

Üniversiteler Mahallesi Dumlupınar Bulvarı No:1

ODTÜ Enformatik Enstitüsü B-207

Ankara, 06800

Turkey

Phone: 90 312 210 7708

Fax: 90 312 210 3745

Email: yesim@metu.edu.tr

Abstract

Background: Personalized medicine approaches provide opportunities for predictive and preventive medicine. Using genomic, clinical, environmental, and behavioral data, tracking and management of individual wellness is possible. A prolific way to carry this personalized approach into routine practices can be accomplished by integrating clinical interpretations of genomic variations into electronic medical records (EMRs)/electronic health records (EHRs). Today, various central EHR infrastructures have been constituted in many countries of the world including Turkey.

Objective: The objective of this study was to concentrate on incorporating the personal single nucleotide polymorphism (SNP) data into the National Health Information System of Turkey (NHIS-T) for disease risk assessment, and evaluate the performance of various predictive models for prostate cancer cases. We present our work as a miniseries containing three parts: (1) an overview of requirements, (2) the incorporation of SNP into the NHIS-T, and (3) an evaluation of SNP incorporated NHIS-T for prostate cancer.

Methods: For the first article of this miniseries, the scientific literature is reviewed and the requirements of SNP data integration into EMRs/EHRs are extracted and presented.

Results: In the literature, basic requirements of genomic-enabled EMRs/EHRs are listed as incorporating genotype data and its clinical interpretation into EMRs/EHRs, developing accurate and accessible clinicogenomic interpretation resources (knowledge bases), interpreting and reinterpreting of variant data, and immersing of clinicogenomic information into the medical decision processes. In this section, we have analyzed these requirements under the subtitles of terminology standards, interoperability standards, clinicogenomic knowledge bases, defining clinical significance, and clinicogenomic decision support.

Conclusions: In order to integrate structured genotype and phenotype data into any system, there is a need to determine data components, terminology standards, and identifiers of clinicogenomic information. Also, we need to determine interoperability standards to share information between different information systems of stakeholders, and develop decision support capability to interpret genomic variations based on the knowledge bases via different assessment approaches.

(*JMIR Med Inform* 2014;2(2):e15) doi: [10.2196/medinform.3169](https://doi.org/10.2196/medinform.3169)

KEYWORDS

health information systems; clinical decision support systems; disease risk model; electronic health record; epigenetics; personalized medicine; single nucleotide polymorphism

Introduction

The digital age is revolutionizing the old and historical population-based health care paradigm toward personalized medicine. Traditional medical approaches are not sufficiently predictive and preventive, as they focus on the manifestation of symptoms that often hide risk factors. Determining risk factors allows for prevention through early diagnosis, and provides new opportunities for developing personalized medicine approaches based on patient-centered, predictive, preventive, and effective health care services [1].

Genomic data and its derivatives (transcriptomes, proteomes, metabolomes, etc) are the essential elements of personalized medicine [2,3]. Every individual has almost four million variations in their own genome, when compared to the reference sequence. Genomic variations can range from single nucleotide changes to the gain or loss of whole chromosomes. Single nucleotide polymorphisms (SNPs), where a single nucleotide in the genome alters between individual or paired chromosomes, are about 90% of genomic variants, and some are already validated as important markers in the clinical practice, while others are on the way [4-6].

The rapid developments in next generation sequencing (NGS) technologies have substantially reduced both the cost and the time required to sequence the entire human genome, and it is expected that NGS-based analyses, for example, whole genome sequencing (WGS) and whole exome sequencing (WES), will be available for routine use in health care and prevention of disease by 2020 [7]. Providing genomic data to medical professionals will facilitate clinical decisions based on the individual's genome, and allow tailoring health care services to the patient's specific needs and characteristics [8]. In parallel, direct-to-consumer (DTC) genome-wide profiling tests are being developed to assess individual disease risks for many common polygenic diseases [9]. DTC genomic companies, for example, 23andMe, GenePlanet, and DNA DTC generally perform a gene-chip analysis of SNPs using deoxyribonucleic acid (DNA) extracted from saliva or serum sample [10-12].

In clinical decision processes, genomic variant data can be used for assessing disease risks, predicting susceptibility, early clinical diagnosing, following the course of the disease, targeted screening, and planning treatment regimens [3,13]. A reasonable way to carry this personalized approach into routine for medical practices would be integrating genotype data and its clinical interpretation within the electronic medical records (EMRs)/electronic health records (EHRs) [8,14].

Today, in many developed and developing countries, use of EMRs/EHRs is inevitable for health care providers for reimbursement of services, and to track the quality of the health care provided [15,16]. Recently, several EHR networks have been constituted in many countries of the world, including the National Health Information System of Turkey (NHIS-T) [17]. These EHR systems and networks have high potential for integrating genomic data in health care practices for personalized medicine.

In this work, as an initial attempt to develop a sophisticated infrastructure, we focused to incorporate the personal SNP data into NHIS-T for disease risk assessment, and evaluated the performance of various predictive models for prostate cancer cases. We presented our work as three parts: (1) a literature review for requirements, (2) the incorporation of SNP into the NHIS-T [18], and (3) an evaluation of SNP incorporated into NHIS-T for prostate cancer [19]. In this part, the scientific literature was reviewed, and the requirements were extracted regarding SNP data integrated EMRs/EHRs.

Methods

The informatics pipeline for genome sequencing can be divided into several analytical steps, for example, base calling, alignment, variant analysis, interpretation, and in all levels different file formats are generated [20-22]. Currently, tools and techniques are developed for automated and reliable analysis, but clinical interpretation of variant data is still a major problem [21].

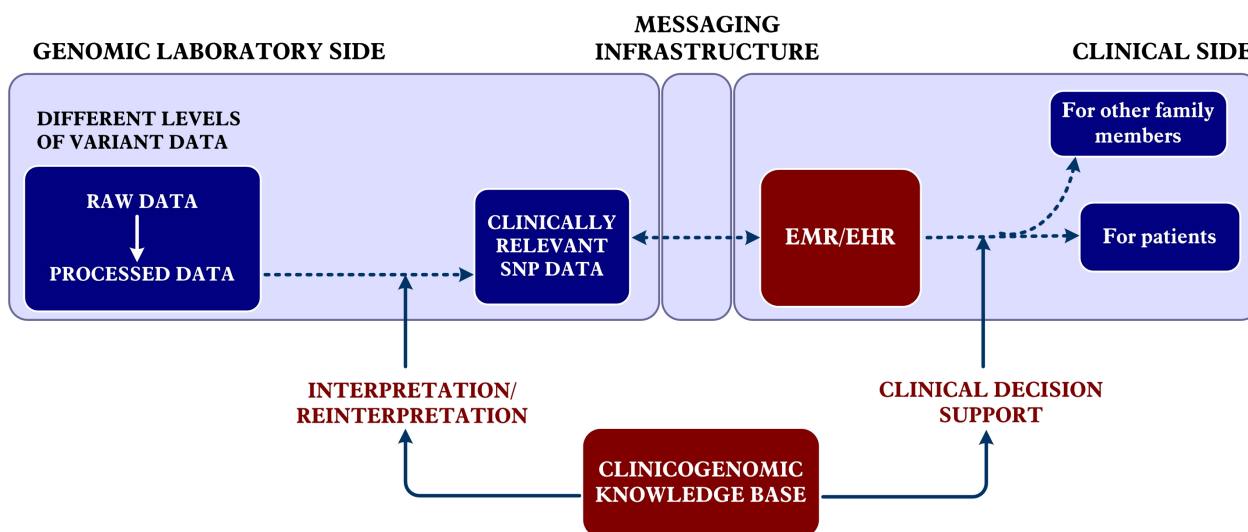
Today, most of the EMRs/EHRs are designed to store and retrieve the laboratory values and clinical findings, but do not have the ability to manage genomic data [23-25]. After WGS/WES, a file that contains a large number of variant data is acquired [26]. An entire genome sequence (the size of the haploid human genome) contains about 3 billion base pairs, and a single WGS data file is about 3 gigabytes. Storing and sharing of personal raw genomic sequences exceeds the transmission and storage capacity in many health care organizations [27]. Due to these technical limitations, raw genomic data are generally stored outside of the EMR; similar to picture archiving and communication systems for medical images, and clinical interpretation of the genomic data is preferably sent to the database of the EMR [28-30].

The initiatives of integrating a patient's genomic data into EMRs/EHRs is of a preliminary nature [31], and, until recently, only a few successful systems have been established, such as Cerner's Genomics Solutions, McKesson's Horizon Clinicals, and GeneInsight [26,32].

In the literature, basic requirements of genomic-enabled EMRs/EHRs are listed as incorporating genotype data and its clinical interpretation into EMRs/EHRs, developing accurate and accessible clinicogenomic interpretation resources (knowledge base), the interpretation and reinterpretation of variant data, and the immersion of clinicogenomic information into the medical decision processes.

Figure 1 shows, in the genome laboratory side, various levels of sequence data can be produced. Since clinicians need an actionable clinical interpretation of the variant data, it is sufficient to share clinically relevant data between the laboratory and the clinical systems. The development of a clinicogenomic knowledge base is an obligation to extract clinical meaning from the variant data. On the clinical side, it is necessary to use decision support systems due to the high number of variants. In some cases, clinicogenomic information may be useful to manage the health status of other family members and other close relatives.

Figure 1. Main components of a genome-enabled electronic medical record/electronic health record. SNP; single nucleotide polymorphisms.



Results

Terminology Standards

In order to integrate structured genotype and phenotype data into any system, the first requirement is to determine data components, terminology standards, and identifiers of clinicogenomic information, for example, genotype data and its associated clinical interpretation.

In genomic terminology, the Human Gene Nomenclature Committee standardizes identifying gene symbols, identifiers, and variant nomenclature defined by the Human Genome Variation Society [6]. Reference SNP number (rs number) and reference SNP identifier (rsID) are used to identify every single SNP entry in the Single Nucleotide Polymorphism Database (dbSNP), which is the largest database maintained by the National Center for Biotechnology Information (NCBI). The dbSNP is interconnected with many other resources, for example, Entrez Gene, GenBank, the Universal Protein Resource, the International HapMap Project, the Pharmacogenomics Knowledge Base (PharmGKB), and the AlzGene, PDGene, SzGene databases through the rsID [33]. Additionally, in many types of personal genomic file formats (eg, 23andMe, deCODEme, and Navigenics), SNPs are identified by rsID.

DNA is a double stranded stretch, and every nucleated somatic cell has 22 pairs of autosomal, and one pair of sex chromosomes. This means for autosomal chromosomes we have two versions of DNA strands inherited via maternal and paternal sex cells. Different forms or variants of a particular polymorphism are called alleles. Because different alleles may have different degrees and types of clinical impact, rsID is insufficient alone to identify the clinicogenomic significance of SNPs. To have a heterozygote allele may not change the risk for the disease, but homozygote allele of the same SNP variant may change the risk for a disease dramatically. For example, in a study, the odds ratio for rs3218536 (A;G) was 0.8 (CI 0.7-1.0), and for the rs3218536 (A;A) 0.3 (0.1-0.9) [34]. Consequently, to identify

clinically relevant SNPs, we need to use a combination of rsID and allele data as the minimum requirements.

DNA has a double strand (plus and minus or forward and reverse stands respectively), and every SNP can be identified using either of the two DNA strands. In various publications, the same alleles of SNPs are defined differently based on the orientation discrepancy [35]. Due to the double-stranded structure of DNA, both approaches are correct, but it is required to declare and use a standard.

Integration of variant data and clinical relevancies bring out the issue of terminological standardization. Unfortunately, conventional health information terminologies do not successfully support the genetic diseases. There is a critical gap between the databases, which involve many terms defining the genetic diseases, and the Systematized Nomenclature of Medicine (SNOMED) [36]. In order to address the chasm between medical vocabularies and bioinformatics resources, the clinical bioinformatics ontology (CBO) was developed and implemented. The CBO is a curated semantic network trying to combine a variety of clinical vocabularies, for example, SNOMED-Clinical Term (CT), Logical Observation Identifiers Names and Codes (LOINC), and NCBI bioinformatics resources [37,38].

In addition, the International Classification of Diseases (ICD) codes, which is also implemented in Turkey, is also preferred for identifying clinical conditions, but the released versions of ICD do not fully support genomic medicine [36]. Existing ICD versions are not efficient to manage all of the levels of clinical, pathologic, and genetic heterogeneities. It is expected that these will be managed in the next version, for example, ICD-11. The ICD-11, which is scheduled for release in 2015, is expected to be interoperable with other medical terminologies such as SNOMED-CT [39]. Nevertheless, it is an unavoidable requirement to develop a new taxonomy of diseases that will be based on information commons and knowledge networks, including a combination of molecular, social, environmental, and clinical data and health outcomes [40].

As explained in the next section, in the clinicogenomic knowledge base, the assessment of both evidence quality of study and effect size of these associations are critical for the analysis of the published results for clinicogenomic associations [41-47]. Despite emerged approaches and initiatives, standardized definitions and value assignment approaches are needed to categorize and use these associations in a consistent way.

Especially for polygenic complex diseases, impact degrees of clinicogenomic association may be different according to race, ethnicity, and environmental factors [48]. The terms of “ethnicity” and “race” refer to a sociocultural construct affecting both biological and environmental factors, and we need a general standard to categorize these terms.

Various predictive models, in clinical settings, may be useful to assess personal disease risk using relevant SNPs, for example, cumulative models, polygenic risk scores, etc. On the other hand, only a small number of holistic enviro-genomic models are available. Because most of the complex diseases are progressing as the interaction of genomic and environmental factors, it seems that, more enviro-genomic models will be produced in near future. Naturally, with the increase of the number and the value of predictive clinicogenomic models, we will need standardized definition and sharing methods for these models.

Interoperability Standards

Health Level 7 (HL7) is a global organization developing health information standards. As an interoperability standard, the HL7 version 2.x (HL7 v2.x) is the most widely used all over the world. HL7 v2.x does not have not a clear information model, and contains many optional data fields. To overcome this vagueness problem, HL7 version 3 (HL7 v3) has been developed, which is based on an object oriented data model called Reference Information Model (RIM) [49]. HL7 v3 Clinical Document Architecture (CDA) is a document markup standard. A HL7 CDA document is produced to exchange information as part of the HL7 v3 standards, and aim to specify the structural and semantic aspects of clinical documents [50].

The HL7 Clinical Genomics (CG) Work Group (WG) develops standards intended to regulate interoperability issues in genomic medicine. *HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model* is based on both the *HL7 Version 2 Implementation Guide Laboratory Result Reporting to the EHR*, and the *HL7 Version 3 Genetic Variation* data model. This guide covers the reporting of the test results for sequencing and genotyping tests, and includes testing for DNA variants associated with diseases and pharmacogenomic applications [36,51,52]. *HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model* was the first example used by The Partners HealthCare Center for Personalized Genetic Medicine and the Intermountain Healthcare Clinical Genetics Institute to gather genetic test results and transmit them to a patient's EHR [51,52]. GeneInsight Suite (GeneInsight Lab, GeneInsight Clinic, and GeneInsight Network) is also a platform where clinical variant data sharing was based on HL7 standards [26,29,53,54].

The HL7 v3 genetic variation specification is based on the HL7 RIM. It uses the HL7 data types, vocabulary binding mechanisms built into the RIM and Bioinformatic Sequence Markup Language to model the sequence information. The root class in the genetic variation model is “genetic loci”, which describes a set of loci, such as a haplotype, a genetic profile, and genetic testing results of multiple variations or gene expression panels. The genetic loci model uses the genetic locus as an information unit to describe each of these loci. A genetic locus is composed of one or more individual alleles, sequences, and observed sequence variations and represents a single gene or coding region. Within this model, HL7 suggests the sharing of the essential part of raw genomic data via “encapsulation”, and extracting clinically relevant data via “bubble-up” based on a genomic decision support application [55].

HL7 CG-WG develops a CDA implementation guide (ie, Implementation Guide for CDA Release 2 Genetic Testing Report) to ensure the transmission of genetic testing reports using HL7 v3 RIM, and is appropriate for the level of granularity of human-readable reports [56].

Clinicogenomic Knowledge Bases

Clinicians cannot extract clinical interpretation of variants directly from the medical sources due to temporal and cognitive limitations [57,58]. So, instead of incorporating all sequence data, integration of the clinical interpretations of variant data into medical records will be more efficient for clinical decision making [54,59]. Therefore, clinically relevant variants must be selected and presented with their clinical meaning, for example, clinicogenomic associations, along with an action plan for clinicians. Since the Human Genome Project, researchers have been discovering new clinicogenomic associations continuously, and it is critical to reinterpret variants and integrate new clinical interpretations into clinical processes [26].

Clinicogenomic associations, which are acquired via studies based on a candidate gene investigation or agnostic screening of complete genome, are published in the scientific literature [41]. Some clinicogenomic knowledge bases collect, curate, interpret, and categorize these published associations between genomic variations and clinical conditions. The Cancer Genome-wide Association and Meta Analyses Database is a part of Cancer Genomic Evidence-based Medicine Knowledge Base, and provides genome-wide association studies (GWAS), research, and meta-analysis about clinicogenomic associations [60,61]. ClinVar provides reports for variations and related phenotypes with evidences [62]. AlzGene [63], PDGene [64], and SzGene [65] are resources, which include manually curated PubMed articles, using systematic methods for Alzheimer's disease, Parkinson's disease, and schizophrenia, respectively. SNPedia is a wiki resource of human genetic variation as published in peer-reviewed research [66]. PharmGKB is a knowledge source containing clinically relevant genotype-phenotype and gene-drug relationships [67].

However, many of the existing knowledge bases for the clinical interpretation of variant data have different conventions. Also, they are not error proof and are not sustainable due to funding issues [54]. Especially for polygenic complex diseases, the impact degrees of clinicogenomic association may be different

according to race, ethnicity, and environmental factors [48]. Therefore, in personalized risk assessment, it will be an ideal approach to use population specific clinicogenomic results, or at least findings from similar communities. If these are not possible, it might be conceivable to use other scientific resources with a confidence range. Experts have been advocating for the generation of centrally curated national repositories of clinically significant variants for the interpretation of an individual's genomic information, eventually [58,68]. To develop a national level clinicogenomic knowledge base is critical to consider consistency of clinicogenomic associations with the sociodemographic characteristics of citizens, and overcome the issues about sustainability.

Regarding published results of clinicogenomic associations, two major points are significant, evidence quality of study and effect size of these associations [41,42]. To measure the magnitude of impact for clinicogenomic associations, researchers usually prefer to use conventional approaches, for example, odds ratio (OR) and relative risks for case control

studies and cohort studies, respectively. These values are presented with CI [43].

In GWAS, many defects and biases might be present based on study design, genotyping, or collected data quality that will affect the clinical value of results [41,44,45]. The quality of evidence is scored based on the type of study and how well the study is conducted [46], and some guidelines are proposed to calculate the evidence degree [47].

Human Genome Epidemiology Network has published the interim Venice guidelines to grade the cumulative evidence in genetic associations. This guideline is based on three criteria: (1) the amount of evidence (sample size), (2) replication of studies (determining association in different studies), and (3) protection from bias (Table 1). After the evaluation of a study, all considerations are categorized as A, B, and C, and finally, merged as a composite assessment using a semiquantitative index as strong, moderate, and weak epidemiological credibility for genetic associations [47].

Table 1. Venice interim guideline criteria for assessment of cumulative evidence on genetic associations [47].

Venice interim guideline criteria	Categories
Amount of evidence	Category A, sample size >1000 Category B, sample size >100 and <1000, Category C, sample size <100 (total number in cases and controls assuming 1:1 ratio)
Extent of replication	Category A, extensive replication including at least one well conducted meta-analysis with little between-study inconsistency. Category B, well conducted meta-analysis with some methodological limitations or moderate between-study inconsistency. Category C, no association; no independent replication; failed replication; scattered studies; flawed meta-analysis; or large inconsistency.
Protection from bias	Category A, bias, if at all present, could affect the magnitude, but probably not the presence of the association. Category B, no obvious bias that may affect the presence of the association, but there is considerable missing information on the generation of evidence. Category C, considerable potential for, or demonstrable bias, which can affect even the presence or absence of the association.

Defining Clinical Significance

Today, Venice criteria are used to assess genomic association studies in several controlled and structured knowledge bases, for example, Alz-Gene, PD-Gene, and SZ-Gene [63-65]. For the importance of clinicogenomic association, some of the knowledge sources include additional data fields that define the magnitude of clinical effects and strength of the relationship between variants and diseases. In ClinVar, clinical significance is defined as a combination of impact and clinical function (eg, benign, pathogenic, protective, drug response, etc), and evidence for clinical significance is categorized regarding study count and type, such as in vitro studies, animal models, etc [62]. In the PharmGKB, a systematic categorization for evidence quality of clinicogenomic associations is extracted depending on methods and results of references [67], but impact value is not emphasized as a parameter. In SNPedia, magnitude is constructed as a subjective measure of interest for magnitude of impact and repute (good, bad) for quality of evidence, but these concepts are not well established. In GET-Evidence,

clinicogenomic references are categorized according to their evidence degree (high, moderate, or low), and clinical significance (high, medium, or low) is used to produce impact score [69].

Clinicogenomic Decision Support

The volume of variation data integrated into clinical practice exceeds the boundaries of unsupported human cognition and interpretive capacity. Additionally, the rapidly growing literature on clinicogenomic associations makes it more complicated to stay current for even experienced professionals [29]. Also, it is not reasonable to expect the interpretation of all clinicogenomic data by the limited number of genetics experts; we need more automated solutions to overcome these obstacles [70]. With the growing data load in the genomic era, in order to make informed decisions in a timely manner, the health care systems need to shift from expert-based practice to systems-supported practice [71].

Although there is a limited number of counter examples, in general, the clinical effect of a single SNP is minor (OR <2.00) [72,73]. Nevertheless, listing of clinicogenomic associations and their effects may be useful to report a limited number of independent associations. This is especially true for disease-associated SNPs with strong impact and strong evidence; users can share these one by one. At this point, using carefully chosen graphics and visualization techniques will be an efficient way of doing so. Various DTC genomic companies report personal genomic risk for various clinical conditions using graphics containing personal estimations [74].

Although the simplest way of reporting SNP variations is displaying these numerous variations in laboratory reports, it is clear that clinicians cannot interpret or evaluate this information stack. Modest value of clinicogenomic associations does not mean negligible, and some researchers try to develop polygenic risk models or panels assigning values for various SNP alleles, and calculate the total risk of disease for more effective risk prediction [75]. In the literature, several cumulative prediction models have been proposed, but most of these are criticized regarding comprehensive evaluation, especially for clinical utility [76].

SNPs could be used to produce a “genomic profile” for disease risk prediction, testing hundreds of thousands of loci across the personal genome. Today, most of the SNP-based risk assessment models have limited predictive utility and discriminative accuracy because most of the disease associated SNPs have small impacts [77,78]. It has been suggested that genomic risk scores based on large numbers of SNPs could explain more about the heritability than models based on a small number and rigorously validated SNPs. But there is a requirement to process large datasets to build such discriminative risk assessment models [79,80].

The genetic architecture of a disease refers to the number, effect size, genetic mode of action (additive, dominant, and/or epistatic), and allelic frequencies of the genetic polymorphisms. The prediction of genetic risk depends on the underlying genetic architecture. Indeed, the SNPs do not have to be the causative mutations. They just need to be in high linkage disequilibrium with the causative mutations so that there is a consistent association between the SNP and disease risk [81].

Different types of polygenic prediction models have been developed to combine the impact of disease associated SNP data, for example, count method, log odds method, multiplicative model, etc. The count method is the calculation of the total count of independent genomic risk alleles. The log odds method sums together the natural logarithm of the allelic OR for each risk allele [78]. DTC testing companies typically employ a multiplicative model to calculate lifetime risk in the absence of an established method for combining SNP risk estimates, for example, multiplication of ORs of each genotype and average population risk [82].

There are various cumulative models combining the impact of several clinicogenomic associations using arithmetic operators.

In recessive models, only homozygote alleles are involved in the models, but in dominant models heterozygote SNPs are also a part of the cumulative models. Both in dominant and recessive models, the values of risk SNPs are accepted as one unit of impact. Models involving alterations of SNPs' impact value regarding homozygote and heterozygote alleles are defined as an additive model [35,43].

Some of the models involve additional criteria, for example, family history [83]. But structured family history is not a mandatory part of EHR, and because of its dynamic characteristics, it is reasonable to collect and trace it at each visit from patients. It is clear that, similar to clinicogenomic associations, collection and reinterpretation of family history is critical to capture effective results with this type of predictive models.

Actually, genomic and environmental factors are involved in various degrees with the molecular etiology of diseases. In monogenetic diseases (eg, Huntington's disease, phenylketonuria, hereditary cancer forms, etc), single gene mutations are predominantly the main cause of diseases. The genetic origins of the complex multifactorial diseases are much more complicated than the monogenetic diseases, which are a result of the complicated interactions between genetic and environmental causes [84].

Genomic information has lifelong value and one's genomic findings can reveal others within families [23]. If a patient is found to have a disease associated variant, possibly other blood relatives would carry the similar risk, and the patient's health care provider could utilize this new clinical information [26]. This is especially important, not only because of the medical perspective, but also for security and privacy issues.

Discussion

In this part of the miniseries, we have reviewed the scientific literature to extract the requirements for SNP data integrated into EMRs/EHRs.

In order to integrate structured genotype and phenotype data into any system, the first requirement is to determine data components, terminology standards, and identifiers of clinicogenomic information, for example, genotype data and its associated clinical interpretation. Also, we need interoperability standards such as HL7 v2 or v3 to share information between stakeholders.

Because of the huge amount of clinically relevant genomic data and fast translation of this information to a clinical domain, we need clinical decision support capability. To ensure this capability, we also need a continuously updated accredited and structured knowledge base, and assessment approaches to interpret these genomic variations.

In the next part of the miniseries, we will present our study to extend capabilities of NHIS-T to handle SNP data, and its clinical interpretation to assess personal disease risk, and propose possible solutions regarding these requirements.

Conflicts of Interest

None declared.

References

1. Downing GJ. Key aspects of health system change on the path to personalized medicine. *Transl Res* 2009 Dec;154(6):272-276. [doi: [10.1016/j.trsl.2009.09.003](https://doi.org/10.1016/j.trsl.2009.09.003)] [Medline: [19931192](https://pubmed.ncbi.nlm.nih.gov/19931192/)]
2. Arnold GL, Vockley J. Thoroughly modern medicine. *Mol Genet Metab* 2011;104(1-2):1-2. [doi: [10.1016/j.ymgme.2011.07.011](https://doi.org/10.1016/j.ymgme.2011.07.011)] [Medline: [21807540](https://pubmed.ncbi.nlm.nih.gov/21807540/)]
3. Ginsburg GS, Willard HF. Genomic and personalized medicine: Foundations and applications. *Transl Res* 2009 Dec;154(6):277-287. [doi: [10.1016/j.trsl.2009.09.005](https://doi.org/10.1016/j.trsl.2009.09.005)] [Medline: [19931193](https://pubmed.ncbi.nlm.nih.gov/19931193/)]
4. Barnes MH. Genetic variation analysis for biomedical researchers: A primer. In: *Genetic variation: Methods and protocols (methods in molecular biology)*. USA: Humana Press; 2010.
5. Drmanac R. Medicine. The ultimate genetic test. *Science* 2012 Jun 1;336(6085):1110-1112. [doi: [10.1126/science.1221037](https://doi.org/10.1126/science.1221037)] [Medline: [22654043](https://pubmed.ncbi.nlm.nih.gov/22654043/)]
6. Poo DC, Cai S, Mah JT. UASIS: Universal automatic SNP identification system. *BMC Genomics* 2011 Nov 30;12 Suppl 3:S9 [FREE Full text] [doi: [10.1186/1471-2164-12-S3-S9](https://doi.org/10.1186/1471-2164-12-S3-S9)] [Medline: [22369494](https://pubmed.ncbi.nlm.nih.gov/22369494/)]
7. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genet Med* 2011 Jun;13(6):499-504. [doi: [10.1097/GIM.0b013e318220aaba](https://doi.org/10.1097/GIM.0b013e318220aaba)] [Medline: [21558861](https://pubmed.ncbi.nlm.nih.gov/21558861/)]
8. Scheuner MT, de Vries H, Kim B, Meili RC, Olmstead SH, Teleki S. Are electronic health records ready for genomic medicine? *Genet Med* 2009 Jul;11(7):510-517. [doi: [10.1097/GIM.0b013e3181a53331](https://doi.org/10.1097/GIM.0b013e3181a53331)] [Medline: [19478682](https://pubmed.ncbi.nlm.nih.gov/19478682/)]
9. Bloss CS, Schork NJ, Topol EJ. Effect of direct-to-consumer genomewide profiling to assess disease risk. *N Engl J Med* 2011 Feb 10;364(6):524-534 [FREE Full text] [doi: [10.1056/NEJMoa1011893](https://doi.org/10.1056/NEJMoa1011893)] [Medline: [21226570](https://pubmed.ncbi.nlm.nih.gov/21226570/)]
10. Helgason A, Stefánsson K. The past, present, and future of direct-to-consumer genetic tests. *Dialogues Clin Neurosci* 2010;12(1):61-68 [FREE Full text] [Medline: [20373667](https://pubmed.ncbi.nlm.nih.gov/20373667/)]
11. Chua EW, Kennedy MA. Current state and future prospects of direct-to-consumer pharmacogenetics. *Front Pharmacol* 2012;3:152 [FREE Full text] [doi: [10.3389/fphar.2012.00152](https://doi.org/10.3389/fphar.2012.00152)] [Medline: [22934000](https://pubmed.ncbi.nlm.nih.gov/22934000/)]
12. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *J Pathol Inform* 2012;3:40 [FREE Full text] [doi: [10.4103/2153-3539.103013](https://doi.org/10.4103/2153-3539.103013)] [Medline: [23248761](https://pubmed.ncbi.nlm.nih.gov/23248761/)]
13. Chan IS, Ginsburg GS. Personalized medicine: Progress and promise. *Annu Rev Genomics Hum Genet* 2011;12:217-244. [doi: [10.1146/annurev-genom-082410-101446](https://doi.org/10.1146/annurev-genom-082410-101446)] [Medline: [21721939](https://pubmed.ncbi.nlm.nih.gov/21721939/)]
14. Belmont J, McGuire AL. The futility of genomic counseling: Essential role of electronic health records. *Genome Med* 2009;1(5):48 [FREE Full text] [doi: [10.1186/gm48](https://doi.org/10.1186/gm48)] [Medline: [19439060](https://pubmed.ncbi.nlm.nih.gov/19439060/)]
15. Garets D, Davis M. Electronic medical records vs electronic health records: Yes, there is a difference. 2006URL: http://www.himssanalytics.org/docs/WP_EMR_EHR.pdf [accessed 2013-12-03] [WebCite Cache ID 6Lb0ZAXxG]
16. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform* 2008 May;77(5):291-304. [doi: [10.1016/j.ijmedinf.2007.09.001](https://doi.org/10.1016/j.ijmedinf.2007.09.001)] [Medline: [17951106](https://pubmed.ncbi.nlm.nih.gov/17951106/)]
17. Healthcare Information and Management Systems Society (HIMSS) Global Enterprise Task Force. Electronic health records: A global perspective, part 1. 2010URL: <http://www.himss.org/files/HIMSSorg/content/files/Globalpt1-edited%20final.pdf> [accessed 2013-12-03] [WebCite Cache ID 6Lb0VPAh0]
18. Beyan T, Aydın Son Y. Incorporation of personal single nucleotide polymorphism (SNP) data into a national level electronic health record for disease risk assessment, part 2: The incorporation of SNP into the National Health Information System of Turkey. *JMIR Med Inform* 2014 (forthcoming). [doi: [10.2196/medinform.3555](https://doi.org/10.2196/medinform.3555)]
19. Beyan T, Aydın Son Y. Incorporation of personal single nucleotide polymorphism (SNP) data into a national level electronic health record for disease risk assessment, part 3: Evaluation for prostate cancer risk assessment. *JMIR Med Inform* 2014 (forthcoming). [doi: [10.2196/medinform.3560](https://doi.org/10.2196/medinform.3560)]
20. Röhm U, Blakeley JA. Data management for high-throughput genomics. 2009 Presented at: CIDR , Fourth Biennial Conference on Innovative Data Systems Research; Asilomar, CA; January , 2009; Asilomar, CA, US p. 4-7 URL: http://www-db.cs.wisc.edu/cidr/cidr2009/Paper_31.pdf
21. Wright C, Burton H, Hall A, Moorithie S, Pokorska-Bocci A, Sagoo G, et al. Next steps in the sequence: The implications of whole genome sequencing for health in the UK. Cambridge, UK: PHG Foundation; 2011.
22. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 2011 Jun;12(6):443-451 [FREE Full text] [doi: [10.1038/nrg2986](https://doi.org/10.1038/nrg2986)] [Medline: [21587300](https://pubmed.ncbi.nlm.nih.gov/21587300/)]
23. Hoffman MA. The genome-enabled electronic medical record. *J Biomed Inform* 2007 Feb;40(1):44-46 [FREE Full text] [doi: [10.1016/j.jbi.2006.02.010](https://doi.org/10.1016/j.jbi.2006.02.010)] [Medline: [16616698](https://pubmed.ncbi.nlm.nih.gov/16616698/)]

24. Sethi P, Theodos K. Translational bioinformatics and healthcare informatics: Computational and ethical challenges. *Perspect Health Inf Manag* 2009;6:1h [[FREE Full text](#)] [Medline: [20169020](#)]
25. Jacob HJ, Abrams K, Bick DP, Brodie K, Dimmock DP, Farrell M, et al. Genomics in clinical practice: Lessons from the front lines. *Sci Transl Med* 2013 Jul 17;5(194):194-195. [doi: [10.1126/scitranslmed.3006468](#)] [Medline: [23863829](#)]
26. Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. *Genet Med* 2012 Apr 5 [[FREE Full text](#)] [doi: [10.1038/gim.2012.19](#)] [Medline: [22481129](#)]
27. Kahn SD. On the future of genomic data. *Science* 2011 Feb 11;331(6018):728-729. [doi: [10.1126/science.1197891](#)] [Medline: [21311016](#)]
28. Starren J, Bottinger E, Dente M, Wood G, Hoffman J. AMIA Summit on Translational Bioinformatics. 2012. Crossing the omic chasm: Integrating omic data into the EHRURL: <http://knowledge.amia.org/amia-55142-tbi2012a-1.649213/t-003-1.649838/f-001-1.649839/a-017-1.649846/an-017-1.649847?qr=1> [accessed 2014-07-18] [[WebCite Cache ID 6R4C6yAtB](#)]
29. Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, et al. Technical desiderata for the integration of genomic data into electronic health records. *J Biomed Inform* 2012 Jun;45(3):419-422 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2011.12.005](#)] [Medline: [22223081](#)]
30. Green RC, Rehm HL, Kohane IS. Clinical genome sequencing. In: Ginsburg GS, Willard HF, editors. *Genomic and personalized medicine*, 2nd ed. Amsterdam, Netherland: Academic Press; 2013:102-122.
31. Jing X, Kay S, Marley T, Hardiker NR, Cimino JJ. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an EHR prototype based on the Continuity of Care Record standard. *J Biomed Inform* 2012 Feb;45(1):82-92 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2011.09.001](#)] [Medline: [21946299](#)]
32. Gerhard GS, Carey DJ, Steele Jr GD. Electronic health records in genomic medicine. In: *Genomic and personalized medicine*, 2nd ed. Amsterdam, Netherland: Academic Press; 2013:287-294.
33. Thomas PE, Klinger R, Furlong LI, Hofmann-Apitius M, Friedrich CM. Challenges in the association of human single nucleotide polymorphism mentions with unique database identifiers. *BMC Bioinformatics* 2011;12 Suppl 4:S4 [[FREE Full text](#)] [doi: [10.1186/1471-2105-12-S4-S4](#)] [Medline: [21992066](#)]
34. Auranen A, Song H, Waterfall C, Dicioccio RA, Kuschel B, Kjaer SK, et al. Polymorphisms in DNA repair genes and epithelial ovarian cancer risk. *Int J Cancer* 2005 Nov 20;117(4):611-618. [doi: [10.1002/ijc.21047](#)] [Medline: [15924337](#)]
35. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: A: Background concepts. *JAMA* 2009 Jan 7;301(1):74-81. [doi: [10.1001/jama.2008.901](#)] [Medline: [19126812](#)]
36. Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. *Hum Mutat* 2011 May;32(5):512-516. [doi: [10.1002/humu.21456](#)] [Medline: [21309042](#)]
37. Hoffman M, Arnoldi C, Chuang I. The clinical bioinformatics ontology: A curated semantic network utilizing RefSeq information. *Pac Symp Biocomput* 2005:139-150. [Medline: [15759621](#)]
38. Hoffman MA, Williams MS. Electronic medical records and personalized medicine. *Hum Genet* 2011 Jul;130(1):33-39. [doi: [10.1007/s00439-011-0992-y](#)] [Medline: [21519832](#)]
39. Zafar A, Ezat WP S. Development of ICD 11: Changes and challenges. *BMC Health Serv Res* 2012;12(Suppl 1):I8. [doi: [10.1186/1472-6963-12-S1-I8](#)]
40. Committee on a Framework for Development a New Taxonomy of Disease, National Research Council. *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. USA: National Academies Press; 2011.
41. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: B: Are the results of the study valid? *JAMA* 2009 Jan 14;301(2):191-197. [doi: [10.1001/jama.2008.946](#)] [Medline: [19141767](#)]
42. Van Allen EM, Wagle N, Levy MA. Clinical analysis and interpretation of cancer genome data. *J Clin Oncol* 2013 May 20;31(15):1825-1833. [doi: [10.1200/JCO.2013.48.7215](#)] [Medline: [23589549](#)]
43. Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? *JAMA* 2009 Jan 21;301(3):304-308. [doi: [10.1001/jama.2008.993](#)] [Medline: [19155457](#)]
44. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008 Mar 19;299(11):1335-1344. [doi: [10.1001/jama.299.11.1335](#)] [Medline: [18349094](#)]
45. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, STrengthening the REporting of Genetic Association Studies. Strengthening the reporting of genetic association studies (STREGA): An extension of the STROBE statement. *PLoS Med* 2009 Feb 3;6(2):e22 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1000022](#)] [Medline: [19192942](#)]
46. Riegelman R. *Public health 101: Healthy people - healthy populations (essential public health)*. USA: Jones & Bartlett Publishers; 2010.
47. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, et al. Assessment of cumulative evidence on genetic associations: Interim guidelines. *Int J Epidemiol* 2008 Feb;37(1):120-132 [[FREE Full text](#)] [doi: [10.1093/ije/dym159](#)] [Medline: [17898028](#)]
48. Stepanov VA. Genomes, populations and diseases: Ethnic genomics and personalized medicine. *Acta Naturae* 2010 Oct;2(4):15-30 [[FREE Full text](#)] [Medline: [22649660](#)]

49. Benson T. Principles of health interoperability HL7 and SNOMED, 2nd ed. New York, US: Springer; 2012:121-141.
50. Boone KW. The CDA™ book. London: Springer-Verlag London Limited; 2011:17-20.
51. Shabo A, Ullman-Cullere M, Pochon P, Huff S, Wood G, McDonald C, et al. HL7 version 2 implementation guide: Clinical genomics; fully loinc-qualified genetic variation model, release 1 (1st informative ballot), HL7 version 2. 2009. URL: http://wiki.hl7.org/images/2/24/V2_CG_LOINC_GENVAR_R1_I2_2009MAY.pdf [accessed 2013-11-29] [WebCite Cache ID 6LUSIAYIi]
52. Ribick A. Health level seven clinical genomics version 2 messaging standard implementation guide successfully transmits genomic data electronically. 2010URL: http://www.hl7.org/documentcenter/public_temp_63D90767-1C23-BA17-0C09452387FC1279/pressreleases/hl7_press_20100119.pdf [accessed 2013-11-29] [WebCite Cache ID 6LURs9kk4]
53. Aronson SJ, Clark EH, Babb LJ, Baxter S, Farwell LM, Funke BH, et al. The GeneInsight Suite: A platform to support laboratory and provider use of DNA-based genetic testing. Hum Mutat 2011 May;32(5):532-536 [FREE Full text] [doi: 10.1002/humu.21470] [Medline: 21432942]
54. Health ROTGRF, Medicine IO. In: Olson S, editor. Integrating large-scale genomic information into clinical practice: Workshop summary. USA: National Academies Press; 2012.
55. Shabo A. Clinical genomics data standards for pharmacogenetics and pharmacogenomics. Pharmacogenomics 2006 Mar;7(2):247-253. [doi: 10.2217/14622416.7.2.247] [Medline: 16515405]
56. Shabo A, Ullman-Cullere MS. Implementation guide for CDA release 2 genetic testing report (GTR) (universal realm) draft standard for trial use September 2012 (developer documentation). 2012. URL: [http://www.hl7.org/documentcenter/public_temp_5765BAE6-1C23-BA17-0C2BC20D21AE785C/wg/clingenomics/docs/Genetic%20Testing%20Report%20\(GTR\)%20-%202012.10.30.pdf](http://www.hl7.org/documentcenter/public_temp_5765BAE6-1C23-BA17-0C2BC20D21AE785C/wg/clingenomics/docs/Genetic%20Testing%20Report%20(GTR)%20-%202012.10.30.pdf) [accessed 2014-04-02] [WebCite Cache ID 6OX7gPgEb]
57. Oetting WS. Clinical genetics & human genome variation: The 2008 Human Genome Variation Society scientific meeting. Hum Mutat 2009 May;30(5):852-856. [doi: 10.1002/humu.20987] [Medline: 19260058]
58. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: A time for omic ancillary systems. JAMA 2013 Mar 27;309(12):1237-1238 [FREE Full text] [doi: 10.1001/jama.2013.1579] [Medline: 23494000]
59. Marian AJ. Medical DNA sequencing. Curr Opin Cardiol 2011 May;26(3):175-180 [FREE Full text] [doi: 10.1097/HCO.0b013e3283459857] [Medline: 21415728]
60. Schully SD, Yu W, McCallum V, Benedicto CB, Dong LM, Wulf A, et al. Cancer GAMAdb: Database of cancer genetic associations from meta-analyses and genome-wide association studies. Eur J Hum Genet 2011 Aug;19(8):928-930 [FREE Full text] [doi: 10.1038/ejhg.2011.53] [Medline: 21487441]
61. Centers for Disease Control and Prevention. Cancer GAMAdb, cancer genomic evidence-based medicine knowledge base. 2013URL: <http://www.hugenavigator.net/CancerGEMKB/caIntegratorStartPage.do> [accessed 2013-11-29] [WebCite Cache ID 6LUQ8RWk4]
62. National Center for Biotechnology Information. ClinVar. 2013. URL: <http://www.ncbi.nlm.nih.gov/clinvar/> [accessed 2013-11-29] [WebCite Cache ID 6LUQcwkqC]
63. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. Nat Genet 2007 Jan;39(1):17-23. [doi: 10.1038/ng1934] [Medline: 17192785]
64. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, 23andMe Genetic Epidemiology of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium, Parkinson's Disease GWAS Consortium, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genet 2012;8(3):e1002548 [FREE Full text] [doi: 10.1371/journal.pgen.1002548] [Medline: 22438815]
65. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. Nat Genet 2008 Jul;40(7):827-834. [doi: 10.1038/ng.171] [Medline: 18583979]
66. Cariaso M, Lennon G. SNPedia: A wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res 2012 Jan;40(Database issue):D1308-D1312 [FREE Full text] [doi: 10.1093/nar/gkr798] [Medline: 22140107]
67. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012 Oct;92(4):414-417 [FREE Full text] [doi: 10.1038/clpt.2012.96] [Medline: 22992668]
68. Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. BMC Med Inform Decis Mak 2009;9:17 [FREE Full text] [doi: 10.1186/1472-6947-9-17] [Medline: 19309514]
69. Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, et al. A public resource facilitating clinical use of genomes. Proc Natl Acad Sci U S A 2012 Jul 24;109(30):11920-11927 [FREE Full text] [doi: 10.1073/pnas.1201904109] [Medline: 22797899]
70. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: A systematic review. J Am Med Inform Assoc 2013;20(2):388-400 [FREE Full text] [doi: 10.1136/amiajnl-2012-000892] [Medline: 22922173]
71. McClellan MB, McGinnis JM, Nabel EG, Olsen LM, Institute of Medicine. Evidence-based medicine and the changing nature of health care: 2007 IOM annual meeting summary. Washington, DC: The National Academies Press; 2008.

72. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 2011 Feb;187(2):367-383 [[FREE Full text](#)] [doi: [10.1534/genetics.110.120907](https://doi.org/10.1534/genetics.110.120907)] [Medline: [21115973](https://pubmed.ncbi.nlm.nih.gov/21115973/)]
73. Kalf RR, Mihaescu R, Kundu S, de Knijff P, Green RC, Janssens AC. Variations in predicted risks in personal genome testing for common complex diseases. *Genet Med* 2014 Jan;16(1):85-91 [[FREE Full text](#)] [doi: [10.1038/gim.2013.80](https://doi.org/10.1038/gim.2013.80)] [Medline: [23807614](https://pubmed.ncbi.nlm.nih.gov/23807614/)]
74. Lautenbach DM, Christensen KD, Sparks JA, Green RC. Communicating genetic risk information for common disorders in the era of genomic medicine. *Annu Rev Genomics Hum Genet* 2013;14:491-513. [doi: [10.1146/annurev-genom-092010-110722](https://doi.org/10.1146/annurev-genom-092010-110722)] [Medline: [24003856](https://pubmed.ncbi.nlm.nih.gov/24003856/)]
75. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010 Jul 8;363(2):166-176. [doi: [10.1056/NEJMra0905980](https://doi.org/10.1056/NEJMra0905980)] [Medline: [20647212](https://pubmed.ncbi.nlm.nih.gov/20647212/)]
76. Little J, Wilson B, Carter R, Walker K, Santaguida P, Tomiak E, et al. Multigene panels in prostate cancer risk assessment, evidence report No. 209. Rockville, MD: AHRQ Publication No.12-E020-EF; 2012 Jul. URL: http://www.effectivehealthcare.ahrq.gov/ehc/products/388/1171/EvidenceReport209_multigenepanels_FinalReport_20120629.pdf[[WebCite Cache ID 6RHGsyhC7](#)]
77. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007 Oct;17(10):1520-1528 [[FREE Full text](#)] [doi: [10.1101/gr.6665407](https://doi.org/10.1101/gr.6665407)] [Medline: [17785532](https://pubmed.ncbi.nlm.nih.gov/17785532/)]
78. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009 Sep 15;18(18):3525-3531 [[FREE Full text](#)] [doi: [10.1093/hmg/ddp295](https://doi.org/10.1093/hmg/ddp295)] [Medline: [19553258](https://pubmed.ncbi.nlm.nih.gov/19553258/)]
79. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet* 2011 Oct 15;20(R2):R182-R188 [[FREE Full text](#)] [doi: [10.1093/hmg/ddr378](https://doi.org/10.1093/hmg/ddr378)] [Medline: [21873261](https://pubmed.ncbi.nlm.nih.gov/21873261/)]
80. Wu J, Pfeiffer RM, Gail MH. Strategies for developing prediction models from genome-wide association studies. *Genet Epidemiol* 2013 Dec;37(8):768-777. [doi: [10.1002/gepi.21762](https://doi.org/10.1002/gepi.21762)] [Medline: [24166696](https://pubmed.ncbi.nlm.nih.gov/24166696/)]
81. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev* 2008 Jun;18(3):257-263. [doi: [10.1016/j.gde.2008.07.006](https://doi.org/10.1016/j.gde.2008.07.006)] [Medline: [18682292](https://pubmed.ncbi.nlm.nih.gov/18682292/)]
82. Nusbaum R, Leventhal KG, Hooker GW, Peshkin BN, Butrick M, Salehzadeh Y, et al. Translational genomic research: Protocol development and initial outcomes following SNP testing for colon cancer risk. *Transl Behav Med* 2013 Mar 1;3(1):17-29 [[FREE Full text](#)] [doi: [10.1007/s13142-012-0149-0](https://doi.org/10.1007/s13142-012-0149-0)] [Medline: [23565131](https://pubmed.ncbi.nlm.nih.gov/23565131/)]
83. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008 Feb 28;358(9):910-919. [doi: [10.1056/NEJMoa075819](https://doi.org/10.1056/NEJMoa075819)] [Medline: [18199855](https://pubmed.ncbi.nlm.nih.gov/18199855/)]
84. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: Advances and prospects. *Hum Mol Genet* 2008 Oct 15;17(R2):R166-R173 [[FREE Full text](#)] [doi: [10.1093/hmg/ddn250](https://doi.org/10.1093/hmg/ddn250)] [Medline: [18852206](https://pubmed.ncbi.nlm.nih.gov/18852206/)]

Abbreviations

- CBO:** clinical bioinformatics ontology
- CDA:** Clinical Document Architecture
- dbSNP:** Single Nucleotide Polymorphism Database
- DNA:** deoxyribonucleic acid
- DTC:** direct-to-consumer
- EHR:** electronic health record
- EMR:** electronic medical record
- GWAS:** genome-wide association studies
- HL7:** Health Level 7
- HL7 v2.x:** HL7 version 2.x
- HL7 v3:** HL7 version 3
- HL7 CG:** Health Level 7 Clinical Genomic
- ICD:** International Classification of Diseases
- LOINC:** Logical Observation Identifiers Names and Codes
- NCBI:** National Center for Biotechnology Information
- NGS:** next generation sequencing
- NHIS-T:** National Health Information System of Turkey
- OR:** odds ratio
- PharmGKB:** Pharmacogenomics Knowledge Base
- RIM:** Reference Information Model
- rsID:** reference SNP identifier
- rs number:** reference SNP number
- SNOMED:** Systematized Nomenclature of Medicine
- SNOMED-CT:** Systematized Nomenclature of Medicine Clinical Term

SNP: single nucleotide polymorphism

WES: whole exome sequencing

WG: Work Group

WGS: whole genome sequencing

Edited by G Eysenbach; submitted 08.12.13; peer-reviewed by W Hammond, A James; comments to author 31.12.13; revised version received 25.05.14; accepted 02.07.14; published 24.07.14

Please cite as:

Beyan T, Aydın Son Y

Incorporation of Personal Single Nucleotide Polymorphism (SNP) Data into a National Level Electronic Health Record for Disease Risk Assessment, Part 1: An Overview of Requirements

JMIR Med Inform 2014;2(2):e15

URL: <http://medinform.jmir.org/2014/2/e15/>

doi: [10.2196/medinform.3169](https://doi.org/10.2196/medinform.3169)

PMID: [25599712](https://pubmed.ncbi.nlm.nih.gov/25599712/)

©Timur Beyan, Ye im Aydın Son. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 24.07.2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.