4-19-2016

A New Framework and Prototype Solution for Clinical Decision Support and Research in Genomics and Other Data-intensive Fields of Medicine

James Evans
University of North Carolina at Chapel Hill, jpevans@med.unc.ed

Kirk Wilhelmsen
University of North Carolina at Chapel Hill, kirk@renci.org

Jonathan Berg
University of North Carolina at Chapel Hill, jonathan_berg@med.unc.edu

Charles Schmitt
University of North Carolina at Chapel Hill, cschmitt@renci.org

See next pages for additional authors

Follow this and additional works at: http://repository.edm-forum.org/egems

Part of the Health Information Technology Commons, and the Other Medicine and Health Sciences Commons

A New Framework and Prototype Solution for Clinical Decision Support and Research in Genomics and Other Data-intensive Fields of Medicine

Abstract

Introduction: In genomics and other fields, it is now possible to capture and store large amounts of data in electronic medical records (EMRs). However, it is not clear if the routine accumulation of massive amounts of (largely uninterpretable) data will yield any health benefits to patients. Nevertheless, the use of large-scale medical data is likely to grow. To meet emerging challenges and facilitate optimal use of genomic data, our institution initiated a comprehensive planning process that addresses the needs of all stakeholders (e.g., patients, families, healthcare providers, researchers, technical staff, administrators). Our experience with this process and a key genomics research project contributed to the proposed framework.

Framework: We propose a two-pronged Genomic Clinical Decision Support System (CDSS) that encompasses the concept of the "*Clinical Mendeliome*" as a patient-centric list of genomic variants that are clinically actionable and introduces the concept of the "*Archival Value Criterion*" as a decision-making formalism that approximates the cost-effectiveness of capturing, storing, and curating genome-scale sequencing data. We describe a prototype Genomic CDSS that we developed as a first step toward implementation of the framework.

Conclusion: The proposed framework and prototype solution are designed to address the perspectives of stakeholders, stimulate effective clinical use of genomic data, drive genomic research, and meet current and future needs. The framework also can be broadly applied to additional fields, including other '-omics' fields. We advocate for the creation of a *Task Force on the Clinical Mendeliome*, charged with defining Clinical Mendeliomes and drafting clinical guidelines for their use.

Disciplines

Health Information Technology | Other Medicine and Health Sciences

Creative Commons License

Authors

James Evans, *University of North Carolina at Chapel Hill*; Kirk Wilhelmsen, *University of North Carolina at Chapel Hill*; Jonathan Berg, *University of North Carolina at Chapel Hill*; Charles Schmitt, *University of North Carolina at Chapel Hill*; Ashok Krishnamurthy, *University of North Carolina at Chapel Hill*; Karamarie Fecho, *RENCI, University of North Carolina at Chapel Hill*; Stanley Ahalt, *University of North Carolina at Chapel Hill*.

eGEMs
Generating Evidence & Methods
to improve patient outcomes

A New Framework and Prototype Solution for Clinical Decision Support and Research in Genomics and Other Data-intensive Fields of Medicine

James P. Evans, MD, PhD;[i,ii] Kirk C. Wilhelmsen, MD, PhD;[i,iii] Jonathan Berg,MD, PhD;[i] Charles P. Schmitt, PhD;[v]
Ashok Krishnamurthy, PhD;[v] Karamarie Fecho, PhD;[v] Stanley C. Ahalt, PhD[iv,v]

ABSTRACT

Introduction: In genomics and other fields, it is now possible to capture and store large amounts of data in electronic medical records (EMRs). However, it is not clear if the routine accumulation of massive amounts of (largely uninterpretable) data will yield any health benefits to patients. Nevertheless, the use of large-scale medical data is likely to grow. To meet emerging challenges and to facilitate optimal use of genomic data, our institution initiated a comprehensive planning process that addresses the needs of all stakeholders (e.g., patients, families, health care providers, researchers, technical staff, administrators). Our experience with this process and a key genomics research project contributed to the proposed framework.

Framework: We propose a two-pronged Genomic Clinical Decision Support System (CDSS) that encompasses the concept of the "*Clinical Mendeliome*" as a patient-centric list of genomic variants that are clinically actionable and introduces the concept of the "*Archival Value Criterion*" as a decision-making formalism that approximates the cost-effectiveness of capturing, storing, and curating genome-scale sequencing data. We describe a prototype Genomic CDSS that we developed as a first step toward implementation of the framework.

Conclusion: The proposed framework and prototype solution are designed to address the perspectives of stakeholders, stimulate effective clinical use of genomic data, drive genomic research, and meet current and future needs. The framework also can be broadly applied to additional fields, including other "-omics" fields. We advocate for the creation of a *Task Force on the Clinical Mendeliome*, charged with defining Clinical Mendeliomes and drafting clinical guidelines for their use.

[i]Department of Genetics, University of North Carolina at Chapel Hill, [ii]Department of Medicine, University of North Carolina at Chapel Hill, [iii]Department of Neurology, University of North Carolina at Chapel Hill, [iv]Department of Computer Science, University of North Carolina at Chapel Hill, [v]Renaissance Computing Institute, University of North Carolina at Chapel Hill

Introduction

In genomics and other data-intensive fields of medicine, such as medical imaging, it is now possible to acquire large amounts of data within a relatively short time frame and at a manageable cost; yet our ability to analyze, interpret, and apply those data to guide clinical care lags far behind. For instance, the GenBank repository currently contains data on an astonishing ~150 billion bases and >150 million sequences, and the number of bases has been doubling roughly every 18 months from 1982 to the present.[1] While the GenBank repository is not restricted to human data, the growth in the data is remarkable and serves as a testament to the pressing need to develop tools to understand and apply those data—both human and nonhuman—to improve the human condition. Indeed, efforts are underway to annotate, archive, and curate genomic data; and publicly available databases such as ClinVar, dbGap, DGV, and the Clinical Genomic Database now contain data on thousands of genes and gene variants, many with extensive annotation on functional impacts and validated phenotypes. Moreover, the National Institute of Health's Clinically Relevant Genetic Variants Resource initiative and the Health Level Seven Clinical Genomics working group are pioneering efforts to harmonize individual efforts and databases. Finally, these issues have been propelled into the national spotlight with President Obama's new Precision Medicine initiative, announced in his 2015 State of the Union address, which aims to pursue whole genome sequencing in a million individuals.

While existing data sources are widely used for research, they are not yet readily applicable for clinical use. Thus, clinical genetic testing typically remains highly targeted, with the majority of genetic tests focused on discrete clinical situations and the relatively small number of genes that exhibit high penetrance and result in recognizable disease when mutated. Those genes for which sequence information clearly provides clinically actionable information (e.g., *BRCA1/2*, *MLH1*, *MSH2*) represent an even more constrained subset.

It remains to be seen whether large-scale data derived from genomics and other data-intensive fields will be widely embraced in clinical medicine. For genomics, Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) could become the most cost-effective and efficient strategies for patient-centered care when any genetic query is clinically indicated. On the other hand, it is not yet clear whether the routine accumulation of massive amounts of (mostly uninterpretable) data from genome-scale sequencing will yield a net benefit in terms of improving health. Indeed, the pursuits of both clinical medicine and public health have traditionally been best served by narrow testing based on substantiated evidence. The application of overly broad testing brings with it many potential problems, including inevitable false positive results, ambiguous results that beg for misinterpretation, and considerable downstream costs no matter how inexpensive the upfront cost of testing is. Nevertheless, genome-scale data and other large-scale medical data are likely to become more common, and the medical enterprise needs to grapple with the central questions of how much data should be stored in the electronic medical record (EMR) and in what format in order to facilitate safe and effective clinical use.

Framework for Genomic Clinical Decision Support

To meet these emerging challenges and facilitate the optimal use of genomic data for patientcentered care and clinical research, we advocate for the development of a two-pronged Clinical Decision Support System (CDSS) for genomics that will: (1) provide the clinician with a dynamic visual snapshot

of only those genomic data that are relevant to an individual patient; and (2) capture, store, and curate more comprehensive genomic data for ready access to address future clinical needs and enable genomic research (Figure 1). Our framework focuses on genomics, as this is a rapidly emerging field of medicine, but we have conceptualized our framework in relation to any data-intensive field of medicine.

Background

The concept of a CDSS dates to a seminal manuscript by Ledley and Lusted[2] that provided a mathematical foundation for the application of computational techniques to improve medical diagnosis. Subsequent work built on that foundation, including an oft-cited study by de Dombal et al.,[3] who conducted a controlled clinical trial demonstrating that a CDSS produced significantly greater accuracy and reliability in the diagnosis of acute abdominal pain than did an attending physician. Numerous CDSS capabilities are now routinely incorporated into a diverse array of medical devices (as safety alerts and reminders) but have yet to be fully integrated into the health care environment.[4] The increasing utilization of EMRs[5] now provides a compelling stimulus for more widespread incorporation of CDSSs for genomics and other data-intensive medical fields into the health care workflow.

Arguments in favor of the development of Genomic CDSSs are not new,[6] and theoretical prototypes have been developed,[7,8] as well as several clinical prototypes.[9–11] The functional and technical challenges also have been thoroughly laid out and include a lack of data standards; difficulties integrating multiple sources of data; absence of consensus in the adjudication of variants; a dearth of standardized phenotype data; incomplete or missing annotation on methodology; engagement of stakeholders; and ethical, legal, and social issues related to privacy, quality control, and

Figure 1. Proposed Framework for Genomic Clinical Decision Support
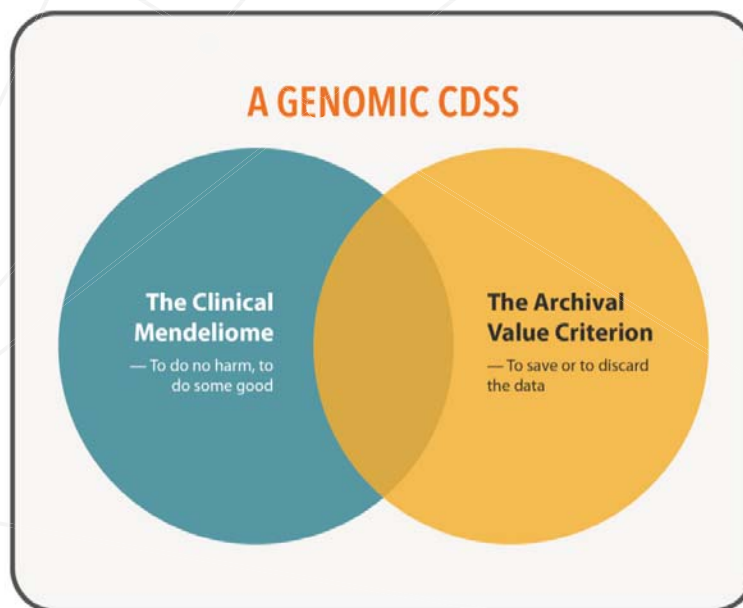
Notes: CDSS = Clinical Decision Support System.

Source: Image courtesy of RENCI.

actionable versus nonactionable information.[12–20] Large-scale efforts such as the nine-institution National Human Genome Research Institute (NHGRI)-funded Electronic Medical Records and Genomics (eMERGE) Network;[21–23] the fifteen-institution NHGRI- and National Cancer Institute (NCI)-funded Implementing Genomics in Practice (IGNITE) Network;[24] and the nine-institution NHGRI- and NCI-funded Clinical Sequencing and Exploratory Research (CSER) Consortium[25,26] are working to resolve these challenges. The eMERGE Network, in particular, has emphasized the need for stakeholder buy-in and Genomic CDSS solutions that are scalable, customizable, transferable, and (largely) automated, and that support an automated architecture that permits the sharing of data, knowledge, and successful clinical interventions.

However, a critical issue that has not received adequate attention is the amount and type of genomic data that should be incorporated into a Genomic CDSS to satisfy the needs of all stakeholders—including clinicians, patients, families, researchers, health care insurance providers, and hospital or medical center administrators and legal representatives. This issue is central to the implementation of not only genomic medicine, but any field of medicine. As with all medical endeavors, the penalty for having too little information is readily apparent—but grave penalties also loom when excessive, redundant, and uninterpretable information populates the EMR, risking pervasive harm through provider distraction and misinterpretation, as well as increased cost.

Historically, CDSSs have been viewed as tools for the clinician to use to improve patient care.

This perspective likely arose and persisted as a result of the initial use of CDSSs as computerassisted diagnostic tools. We argue that this clinician-centric focus, while valuable, should be expanded from providing mere passive alerts for the clinician to being a proactive model that encompasses the perspectives of the diverse stakeholders who participate in both patient care and research. Such a change in framework can in turn guide decisions about which data should be incorporated into the EMR and associated CDSS. A balance must be reached between incorporating too little information and too much information in order to ensure that the EMR contains only meaningful, easy-to-interpret, clinically actionable information.

Methodology

Several key activities contributed to the conceptualization of the proposed genomic CDSS framework. First, recognizing the numerous technical, ethical, legal, and social challenges presented by whole-genome sequencing, as well as its eventual routine use, our institution—the University of North Carolina at Chapel Hill (UNC-CH)—established a formal process to create policies and procedures to direct the use of genomic data for clinical purposes and research. The process was designed to protect the interests of all stakeholders: patients, families, health care providers, researchers, technical support teams, and university administrators.

Three committees were established to initiate the process, as follows.

1. *A steering committee* composed of high-level university administrators—deans and directors from Lineberger Comprehensive Cancer Center, Renaissance Computing Institute (RENCI), the North Carolina Translational and Clinical Sciences Institute (funded through an NIH Center for Translational Science Award), and Information Technology Services Research Computing—was tasked with administrative and legal considerations, including access rights and cost-effectiveness.

2. *A faculty committee* consisting of clinical genomics faculty members (Department of Genetics and Carolina Center for Genome Science) was charged with best practices for clinical and research applications of genomics data, including ethical issues.

3. *A technical committee* composed of technical experts and support staff derived from several academic units at UNC-CH—RENCI, Information Technology Services Research Computing, the North Carolina Translational and Clinical Sciences Institute, and the Lineberger Comprehensive Cancer Center—was tasked with technical considerations, including security.

These committees met both separately and jointly over the course of a year. The outcome of the committee meetings was the formation of the Genomics Task Force and a strategic planning process. The task force comprised select members derived from each committee and was charged with developing a strategic plan and recommendations to execute that plan.

The second key activity that contributed to our proposed framework is our experience with a genomic sequencing project—North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing (NCGENES). This project, which is housed in the Department of Genetics at UNC-CH and is funded as part of the CSER initiative of the National Human Genome Resource Institute, was initiated concurrently with the Genomics Task Force. NCGENES aims to evaluate the use of WES as a diagnostic tool for practical clinical use—in terms of diagnostic performance and implications for patients and families—and to establish a set of best practices to guide future implementation of genomic technologies to improve patient care. NCGENES has sequenced, to date, over 600 patient samples, focusing on those patients in whom there is reason to suspect that a genetic mutation underlies the

patient's disease, including, for example, patients with a strong family history of cancer or pediatric patients with certain developmental disorders. An additional, but critically important, goal of NCGENES is to analyze the clinical utility and actionability and the ethical impact of incidental findings, or genetic variants that may or may not be clinically relevant but are identified as part of WES.

The multidisciplinary NCGENES team worked with the Genomics Task Force to develop a strategic plan for genomics and recommendations to achieve that plan. From the outset, the NCGENES team encountered challenges to the execution of the proposed work.[27–30] Technical challenges included the size of the data sets, the need for reference data sets that are continuously updated, legacy systems, network idiosyncrasies, distributed and uncoordinated compute resources, and diverse and evolving computational workflow needs. Perhaps more significant, however, were the complex sociological, psychological, ethical, cultural, and political challenges that arose during the initiation of NCGENES; these included political and cultural resistance to change, human roadblocks to the automation of workflow tasks, distributed decision-making, weak or broken communication channels, and administrative and legal concerns related to the privacy and security of patient data. The NCGENES team worked with the Genomics Task Force and numerous technical experts in the affiliated academic units to overcome these challenges and implement the recommended policies and procedures. The result was the development and implementation of several sophisticated genomic technologies that together provide a prototype Genomic CDSS to support NCGENES and several other clinical genomics projects at our institution. The features of the prototype NCGENES CDSS are described below, in the "Driving Use Case: NCGENES" section of this paper.

Our experience with NCGENES, as well as the findings of the Genomic Task Force, led to the development of the conceptual framework presented herein and described in detail below. We envision a two-pronged CDSS designed to enable optimal use of genomics data. While developed for genomics, the envisioned CDSS has application to other data-intensive fields of medicine, as described in the concluding section.

The Clinical Mendeliome

For the first prong, we advocate the concept of the EMR retaining information focused upon the "*Clinical Mendeliome*," which we envision as "a patient-centric list of genomic variants that potentially affect clinical care, guided by the principals of 'do no harm' and 'do some good.'" We use the term Clinical Mendeliome not to suggest that all genetic diseases are Mendelian, but rather to capture the essence and simplicity of our approach, which targets only those variants with established clinical validity and utility. This approach addresses the dilemma of both too little and too much information—by first prioritizing variants found in genes known to be implicated in human disease (currently ~3,000 of ~22,000 genes). The variants to be included would be based upon evidence from published literature and public databases and would emphasize a minimal set of genetic variants of known status with regard to utility, pathogenicity, and actionable treatment options. Importantly, these variants would be linked in the Genomic CDSS to pertinent aspects of the patient's demographics and history. Such a panel of variants in a select number of genes, the Clinical Mendeliome, could be automatically generated through a rules-based search engine or selected on the basis of a defined number of parameters chosen by clinicians (i.e., a clinical "order").

The relevance of genomic variation for any given individual is critically affected by context. Thus, an individual patient's clinical status and demographics must guide the construction of variant lists in the Genomic CDSS such that those variants most relevant to a patient's care, in the appropriate context, would be presented for human analysis (see the "Driving Use Case: NCGENES" section of this paper for a prototype solution.) In healthy individuals (i.e., those without obvious disease), for whom genomic analysis is pursued as a screening modality for prevention,[31] a highly curated gene list based upon general demographics would be generated, and only those variants meeting a very high bar with regard to pathogenicity, predictive value, clinical utility, or pharmacogenomic relevance would be displayed in the Genomic CDSS. In each context, the Genomic CDSS must include a simple, dashboard-style interface that would display the Mendeliome results for individual patients, along with complete annotation on each variant (including a measure of reliability of the variant) and an algorithm of actionable clinical measures that should be taken, again defined using evidence from the published literature and publicly available databases. One could envision an extension of this interface to include a visual representation of variant data and treatment outcomes for a comparative population of patients with a similar genetic profile.[32] The Mendeliome results could also be incorporated into a patient's Personal Health Record,[33] in line with the Meaningful Use incentives afforded by the Health Information Technology for Economic and Clinical Health Act of 2009,[34] as well as the transition from volume- to value-driven accountable care, and satisfying the long-term interests of patients, clinicians, and health care insurance providers.

The Archival Value Criterion

For the second prong, we introduce the concept of the "*Archival Value Criterion*," which is "a decision-making formalism that addresses the dilemma of whether 'to save' or 'to discard' genomic data and

other large-scale medical data." We envision that WGS and WES data will be incorporated into the EMR as an underlying data layer—perhaps as Variant Call Format (vcf) files—that includes extensive annotation on how the data were collected, stored, and curated and a reliability or confidence measure on the quality of the data. These data would not be proximal to the clinician, so as to eliminate confounding and irrelevant data, but the data would be accessible and available for both future clinical needs (as new genomic information arises and the Mendeliome is updated accordingly) and research purposes (with appropriate consent or deidentification[35,36]). The data would also be available for upgrade to the clinician dashboard, allowing proactive monitoring and alerting for cases in which new annotation or pharmacogenomic indicators arise. The data could also be combined with (de-identified) data from the Genomic CDSS, thereby enabling clinical care to directly stimulate and guide new research.

The cost-effectiveness of capturing, storing, and curating WGS- and WES-derived data is simple to calculate to a first approximation. We suggest the following formalism as the Archival Value Criterion (AVC) for determining cost-effectiveness: $AVC = (P_{reuse} \times S')/S$, where S is the total cost for the storage and curation of sequencing data, P_{reuse} is the estimated probability of reuse, and S' is the cost of re-generation. We propose that an AVC metric $>10^2$ suggests data archiving rather than data re-generation. After data have been archived, the AVC metric for removal should be much lower, perhaps $<10^{-6}$. We note that the AVC is influenced by numerous factors, including compression, secondary factors such as staff time and resource reallocation, and timedependent parameters (see Wilhelmsen et al.[37] for a detailed discussion), but the AVC as proposed provides a useful formalism to estimate the cost-effectiveness of archiving genomic data.

Driving Use Case: NCGENES

As discussed in the "Methods" section of this paper, the proposed framework was developed, to a large extent, in response to the challenges our team faced when attempting to implement a large-scale, federally funded clinical genomics project, NCGENES. NCGENES is complex and has both clinical and research arms. The primary aim of NCGENES is to evaluate the use and impact of both diagnostic and incidental findings obtained as a result of WGS and WES and develop practical and ethical mechanisms to facilitate interpretation, management, reporting, and decision-making by both clinicians and patients. NCGENES also aims to use the genetic information obtained in the course of the diagnostic evaluation as a resource to conduct research into the causes of various genetic disorders. A secondary aim of NCGENES is to serve as a prototype for a Genomic CDSS as envisioned by the authors of this paper.

For NCGENES, WES results are analyzed and genomic variants are categorized via a "binning" system that the team developed. The binning system classifies genomic variants according to their clinical utility and validity: clinically relevant diagnostic results (i.e., genes suspected of causing the patient's disease); medically actionable incidental findings; and medically nonactionable incidental findings. The medically actionable incidental findings are subclassified as having both clinical utility and validity (e.g., *BRCA1/2*, *MLH1*, *MSH2*, *FBN1*, *NF1*) or having clinical validity but questionable clinical utility; the latter were subclassified as low risk (e.g., PGx variants, common risk SNPs), medium risk (e.g., *APOE*, carrier status for recessive Mendelian disorders), or high risk (e.g., Huntington's Disease, Prion Diseases, ALS). Variants determined to be clinically relevant diagnostic results are reported to all study subjects; variants determined to be

medically actionable incidental findings are reported only to study subjects who consent in advance to receive that information; medically nonactionable incidental findings are used for research purposes only, in an effort to uncover new genetic causes of disease. This approach thus empowers the patient in his or her decision-making. *Importantly, the binning system also sparked the* concept of the Clinical Mendeliome *as presented herein*.

The challenges to the implementation of NCGENES, in general, and the sequencing and binning system, in particular, were numerous and involved technical, sociological, psychological, ethical, cultural, and political issues.[27-30] Technical challenges involved the size of the sequencing data sets, the need to provision reference data and continuously update those data, integration with legacy systems, network idiosyncrasies, distributed and uncoordinated compute resources, and diverse and evolving computational workflow needs. Nontechnical challenges involved political and cultural resistance to the automation of workflow tasks, an infrastructure based on distributed decision-making, weak or broken communication channels, and administrative and legal concerns related to the privacy and security of patient data.

The NCGENES team worked closely with administrative, legal, and technical teams to overcome these challenges—largely through the development and deployment of three, homegrown key technologies: (1) the GMW (Genetic Medical Workflow) Engine;[28] (2) MaPSeq (Massively Parallel Sequencing);[29,30] and (3) the CANVAS (CAroliNa Variant Annotation Store) and AnnoBot (Annotation Bot) system.[27] Several other technologies are also integral to NCGENES but were developed prior to NCGENES; these include the integrated Rule-Oriented Data System (iRODS)[38-41] and the Secure Medical Workspace (SMW).[42,43]

Figure 2 provides a high-level overview of NCGENES operations in terms of the underlying informatics infrastructure, overall operations, and roles of various personnel. The GMW Engine largely drives the entire project and enables clinical and patient decision-making by providing secure, end-to-end capture, analysis, validation, and reporting of WES for both research and clinical care. The GMW Engine accomplishes these tasks by managing and orchestrating all people, processes, samples, and information involved in the NCGENES project. Security is provided by a number of available options, including the use of the SMW to provision a virtual environment for access to all data that contain protected health information (PHI) and iRODS as a comprehensive, policy-based, secure data management system. Clinically actionable genomic findings and all PHI data are stored within our Epic EMR system; the full set of sequencing results and variant annotations, stripped free of PHI, are stored on researcher-funded equipment in order to remove the storage burden from the institution but allow the data to be accessible for future research and clinical use (the latter by way of linkage identifiers between patient sequencing data and the patient EMR). *This division of storage burden and the consequential financial burden that such a division placed upon researchers largely inspired our AVC as a rational mathematical formalism of the cost-effectiveness of archival of WGS and WES data.*

Figure 3 provides a more detailed view of the GMW Engine, which drives the overall system. The GMW Engine is a highly sophisticated and powerful workflow engine. While much of the workflow processes and technologies, such as the SMW and iRODS, were already in place at our institution before NCGENES was implemented, they lacked coordination and the sophistication needed to bring NCGENES to fruition. The GMW Engine provides a

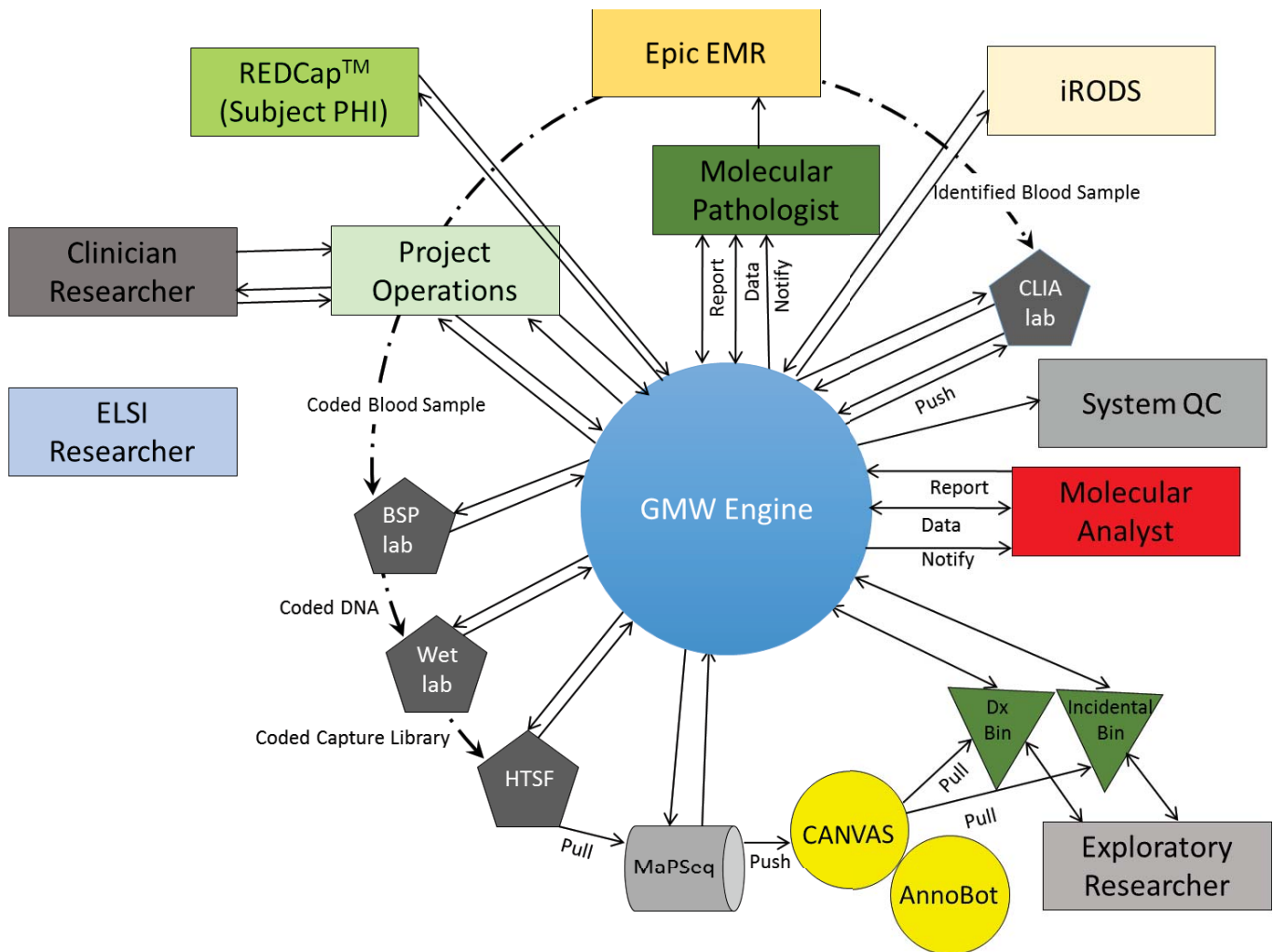## Figure 2. Overview of NCGENES Operations and Capabilities
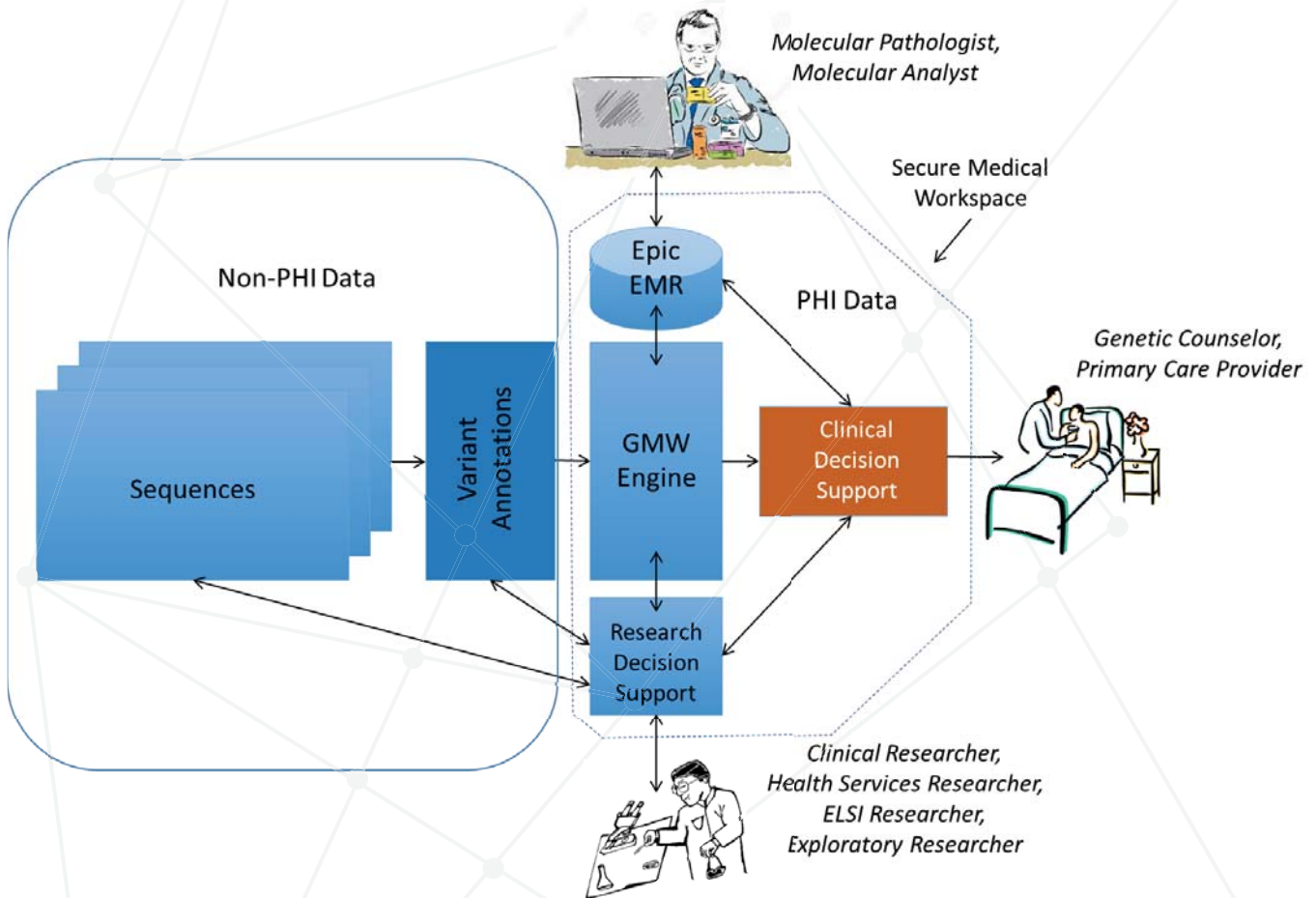
Notes: ELSI Researcher = Ethical, Legal, and Social Implications Researcher; EMR = electronic medical record; GMW Engine = Genetic Medical Workflow Engine; PHI = protected health information.

Source: Image courtesy of RENCI

seamless integration and orchestration of disparate and distributed systems, processes, samples, data, and people. In addition to the GMW Engine, the NCGENES team developed and implemented two new technologies that have since proven to be fundamental to the success of the project: MaPSeq and CANVAS/AnnoBot. MaPSeq securely manages and executes the complex downstream computational and analytical workflow steps required for genomic sequencing, including the opportunistic use of distributed compute resources. CANVAS and AnnoBot together provide version-controlled annotation and metadata on genomic variant data, with continuous, automated updates from monitored databases and national repositories. While each of these technologies is being refined over time, all have proven effective and efficient in the implementation of NCGENES.

Figure 3. Schematic Showing the Workflows Managed by the GMW Engine

Notes: Arrows depict the information flow. AnnoBot = Annotation Bot; BSP lab = BioSpecimen Processing laboratory; CANVAS = CAroliNa Variant Annotation Store; CLIA lab = a laboratory certified to meet United States Congressional Clinical Laboratory Improvements Amendments; Dx = diagnostic; ELSI Researcher = Ethical, Legal, and Social Implications Researcher; EMR = electronic medical record; GMW Engine = Genetic Medical Workflow Engine; iRODS = integrated Rule-Oriented Data System; MaPSeq = Massively Parallel Sequencing system; PHI = protected health information; QC = Quality Control; Wet lab = basic science laboratory.

Source: Image reprinted with permission from Owen et al.[28]

Figure 4 presents a screenshot of the clinician interface for the NCGENES CDSS. The interface provides a comprehensive snapshot of a given patient's status, in terms of both the NCGENES study and the clinical genomic findings. The interface remains in development and is undergoing continual evaluation but has shown promise for both improving decision-making on the part of both clinicians and their patients. Figure 5 shows a prototype of a more advanced NCGENES CDSS interface that we've termed "Carnac" (in homage of Johnny Carson's Carnac the Magnificent comedic role). This interface attempts to capture a more patient-centered view of relevant genomic variants, in the context of the implicated gene and genetic locus (the cardiomyopathy-associated *AGL* gene in this example; Figure 5A), the American College of Medical Genetics current body of evidence regarding the identified genomic variant (Figure 5B), and a comparative, population-level overview of all alleles at that genetic locus (left side of Figures 5A and 5B). The clinician interface is but one of numerous

Figure 4. NCGENES CDSS User Interface Showing Study Status and Results for a Study Participant

Notes: Dx = Diagnostic; ID = identifier.

Source: Image courtesy of NCGENES.

Figure 5. Prototype NCGENES CDSS User Interface (Termed "Carnac")

Notes: Carnac provides detailed views of the implicated gene and genetic locus (panel A); the American College of Medical Genetics current body of evidence regarding the identified genomic variant (panel B); and a comparative, population-level overview of all alleles at that genetic locus (left side of panels A and B).

Source: Images courtesy of NCGENES.

interfaces that are available to NCGENES team members; each interface is tailored to a particular role and presents only that set of information that is necessary to a person's specific role (e.g., molecular analyst, molecular pathologist). A goal of the project is to simplify the user interfaces, as they currently entail significant training in their use.

In addition to ongoing research and development of the technologies supporting the NCGENES project and the prototype NCGENES CDSS, we have successfully adapted our approach to support several other federally funded, clinical genomics projects, including the NICHD–funded North Carolina Newborn Exome Sequencing and Newborn Screening Disorders (NC Nexus), which aims to conduct whole exome sequencing on 400 patient samples; UNCSeq, which applies tumor sequencing technology for >2,000 patient samples in order to identify mutations that are amenable to targeted treatments; and the National Institute on Drug Abuse (NIDA)-funded NIDASeq or Deep Sequencing Studies for Cannabis and Stimulant Dependence, which is conducting whole genome sequencing of ~5,500 patient samples. In addition, the GMW Engine aggregates and stores ~6,000 additional genomes that are derived from public databases and are used for analysis in ongoing genomic research studies; these are obtained from the 1000 Genomes project, The Cancer Genome Atlas project, the national Exome Sequencing Project, and Complete Genomics. Moreover, we note that the GMW Engine and CANVAS/Annobot support the NIH-funded ClinGen initiative, which involves a national effort to develop consensus annotation for the NIH ClinVar database.

Conclusion and Next Steps

Our vision of a two-pronged Genomic CDSS encompasses the concept of the Clinical Mendeliome and introduces the concept of the AVC. Our model builds upon our experience with the NCGENES project and related endeavors. The model embraces the perspectives and facilitates the goals of stakeholders from proximal clinical encounter to distal research endeavors. Such an approach enables genomic medicine to accelerate genomics research and vice versa, and it provides incentives to promote industry commoditization of the stored genomic data.

The proposed framework was developed specifically for genomics, but it has widespread application across many fields of medicine and various types of research such as patient-centered outcomes research and delivery system science. For example, in other work, we developed a visualization-based CDSS that compares defined treatment outcomes in individual patients with major depressive disorder with those of a comparative population, drawn from the EMRs of three collaborating institutions.[32] Our genomic CDSS could be integrated with this system to enable comparative effectiveness research and clinical decision support on genomic factors influencing treatment outcomes in patients with major depressive disorder. To provide another example, consider a patient-centered outcomes study in which patients with early-stage lung cancer self-report daily symptoms (type, severity) while undergoing an experimental treatment. Our genomic CDSS could be adapted to incorporate the self-reported symptom data in order to identify genomic factors influencing response to treatment. Alternatively, if imaging data are regularly collected as part of the study, then our genomic CDSS could be modified to capture those data and present only the most salient information (i.e., tumor imaging) to the health care provider in order to examine the correlation between genomic variants, self-reported symptoms, and tumor response to treatment; additionally, both the actionable and less actionable imaging data (i.e., imaging of secondary structures) could be archived for secondary research purposes.

These are just a few examples. The clinical application of data derived from any research design or medical field must be guided by the identification of clinically actionable findings for any given patient, and one can thus envision the creation of personalized "Clinical Proteomes," "Clinical Metabolomes," etc. The AVC, as presented here, is field-agnostic and can be applied to any data-intensive field of medicine. Moreover, our general CDSS framework can be extended to integrate medical data across different domains, thus enabling truly comprehensive clinical decision support and research.

As a first step toward realizing our vision and implementing our Genomic CDSS model on a national scale, we recommend the creation of a *Task Force on the Clinical Mendeliome*. The task force would be charged with defining context-specific Clinical Mendeliomes and creating an initial set of clinical practice guidelines for their use. The task force would comprise various stakeholders with expertise in clinical genomics, genetics, information technology, computer science, and bioethics— and would include, for example, members of the American College of Medical Genetics and the National Consortium for Data Science, as well as members representing the Clinically Relevant Genetic Variants Resource initiative and the Health Level Seven Clinical Genomics working group. Patients, their families, and their providers represent key stakeholders who also should be included in the task force.[16] Indeed, the input of patients, families, and providers is critical to ensure that the interests of researchers and information technology staff, as well as other "outside" groups, remain aligned with that of the individuals at the center of the proposed framework. By bringing together stakeholders with different interests and areas of expertise, the task force would be well positioned to identify solutions to the inevitable challenges that will arise when addressing the central question of how much genomic data should be captured, stored, and curated for the care of patients and furtherance of research.

References

1. GenBank, www.ncbi.nlm.nih.gov/genbank/statistics.
2. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. Science. 1959;130(3366):9–21.
3. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. Br Med J. 1972;2(5804):9–13.
4. Mitchell JA, Gerdin U, Lindberg DA, Lovis C, Martin-Sanchez FJ, Miller RA, Shorliffe EH, Leong TY. 50 years of informatics research on decision support: what's next. Methods Inf Med. 2011;50:525–535.
5. Charles D, Gabriel M, Furukawa MF. Adoption of Electronic Health Record systems among U.S. non-federal acute care hospitals: 2008–2013; 2014 May. ONC Data Brief No. 16. www.healthit.gov/sites/default/files/oncdatabrief16.pdf.
6. Hoffman MA. The genome-enabled electronic medical record. J Biomed Inform. 2007;40(1):44–46.
7. Douali N, Jaulent MC. Genomic and personalized medicine decision support system. IEEE 2012 International Conference on Complex Systems (ICCS), 2012. Nov 5-6; Agadir, Morocco. IEEE; 2012. pp. 1–4. doi: 10.1109/ICoCS.2012.6458611.
8. Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, Hoffman MA, Kohane IS, Levy HP. Technical desiderata for the integration of genomic data into electronic health records. J Biomed Inform. 2012;45(3):419–422.
9. Neri PM, Pollard SE, Volk LA, Newmark LP, Varugheese M, Baxter S, Aronson SJ, Rehm HL, Bates DW. Usability of a novel clinician interface for genetic results. J Biomed Inform. 2012;45:950–957.
10. Foreman AK, Berg J, Roche M, et al. A year of NCGENES: North Carolina Clincial Genomic Evaluation by NextGen Exome Sequencing [abstract]. Presented at the American College of Medical Genetics Annual Clinical Genetics Meeting; 2013 March 21; Phoenix, AZ. ww2.aievolution.com/acm1301/index.cfm?do=abs.viewAbs&abs=1570.
11. Welch BM, Salvador RL, Eilbeck K, Kawamoto K. Clinical decision support for whole genome sequence information leveraging a service-oriented architecture: a prototype. AMIA Annu Symp Proc. 2014;2014:1188–1197. eCollection 2014. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419907/.
12. 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. www.biomedcentral.com/1472-6947/9/17.
13. Ahalt S, Bizon C, Evans J, Erlich Y, Ginsberg G, Krishnamurthy A, Lange L, Maltbie D, Masys D, Schmitt C, Wilhelmsen K. Data to discovery: genomes to Health. A white paper from the National Consortium for Data Science. RENCI/National Consortium for Data Science White Paper. RENCI, University of North Carolina at Chapel Hill, 2013. doi: 10.7921/G03X84K4. data2discovery.org/publications.
14. Jacob HJ, Abrams K, Bick DP, Brodie K, Dimmock DP, Farrell M, Geurts J, Harris J, Helbling D, Joers BJ, Kliegman R, Kowalski G, Lazar J, Margolis DA, North P, Northup J, Roquemore-Goins A, Scharer G, Shimoyama M, Strong K, Taylor B, Tsaih SW, Tschannen MR, Veith RL, Wendt-Andrae J, Wilk B, Worthey EA.. Genomics in clinical practice: lessons from the front lines. Sci Transl Med. 2013;5(194):194cm5. stm.sciencemag.org/content/5/194/194cm5.abstract.
15. Marsolo K, Spooner SA. Clinical genomics in the world of the electronic health record. Genet Med. 2013;15(10):786–791.
16. Hartzler A, McCarty CA, Rasmussen LV, Williams MS, Brilliant M, Bowton EA, Clayton EW, Faucett WA, Ferryman K, Field JR, Fullerton SM, Horowitz CR, Koenig BA, McCormick JB, Ralston JD, Sanderson SC, Smith ME, Trinidad SB. Stakeholder engagement: a key component of integrating genomic information into electronic health records. Genet Med. 2013;15(1):792–801. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3909653/.
17. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, Williams MS, Clayton EW, Kullo IJ. Ethical, legal, and social implications of incorporating genomic information into electronic health records. Genet Med. 2013;15(10):810–816. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3926430/.
18. Welch BM, Eilbeck K, Del Fiol G, Meyer LJ, Kawamotor K. Technical desiderata for the integration of genomic data with clinical decision support. J Biomed Inform. 2014;51:3–7.
19. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, Goy A, Suh KS. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma. 2015;5:4. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4381462/.
20. Lewis MA, Paquin RS, Roche MI, Furberg RD, Rini C, Berg JS, Powell CM, Bailey DB. Supporting parental decisions about genomic sequencing for newborn screening: the NC NEXUS decision aid. Pediatrics. 2016;137(s1):S16–S23.
21. Overby CL, Kohane I, Kannry JL, Williams MS, Starren J, Botinger E, Gottesman O, Denny JC, Weng C, Tarczy-Hornoch P, Hripcsak G. Opportunities for genomic clinical decision support interventions. Genet Med. 2013;15(10):817–823.
22. Crawford DC, Crosslin DR, Tromp G, Kullo IJ, Kuivaniemi H, Hayes MG, Denny JC, Bush WS, Haines JL, Roden DM, McCarty CA, Jarvik GP, Ritchie MD. eMERGEing progress in genomics – the first seven years. Front Genet. 2014;5:184. eCollection 2014. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4060012/.
23. Herr TM, Bielinski SJ, Bottinger E, Brautbar A, Brilliant M, Chute CG, Cobb BL, Denny JC, Hakonarson H, Hartzler AL, Hripcsak G, Kannry J, Kohane IS, Kullo IJ, Lin S, Manzi S, Marsolo K, Overby CL, Pathak J, Peissig P, Pulley J, Ralston J, Rasmussen L, Roden DM, Tromp G, Uphoff T, Weng C, Wolf W, Williams MS, Starren J. Practical considerations in genomic decision support: the eMERGE experience. J Pathol Inform. 2015;6:50. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4629307/.

24. Weitzel KW, Alexander M, Bernhardt BA, Calman N, Carey DJ, Cavallari LH, Field JR, Hauser D, Junkins HA, Levin PA, Levy K, Madden EB, Manolio TA, Odgis J, Orlando LA, Pyeritz R, Wu RR, Shuldiner AR, Bottinger EP, Denny JC, Dexter PR, Flockhart DA, Horowitz CR, Johnson JA, Kimmel SE, Levy MA, Pollin TI, Ginsburg GS, on behalf of the IGNITE Network. The IGNITE Network: a model for genomic medicine implementation and research. BMC Med Genomics 2015;9:1. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4700677/.

25. Amendola LM, Lautenbach D, Scollon S, Bernhardt B, Biswas S, East K, Everett J, Gilmore MJ, Himes P, Raymond VM, Wynn J, Hart R, CSER Genetic Counseling Working Group, Jarvik GP. Illustrative case studies in the return of exome and genome sequencing results. Per Med. 2015;12(3):283–295.

26. Shirts BH, Salama JS, Aronson SJ, Chung WK, Gray SW, Hindorff LA, Jarvik GP, Plon SE, Stoffel EM, Tarczy-Hornoch PZ, Van Allen EM, Weck KE, Chute CG, Freimuth RR, Grundmeier RW, Hartzler AL, Li R, Peissig PL, Peterson JF, Rasmussen LV, Starren JB, Williams MS, Overby CL. CSERT and eMERGE: current and potential state of the display of genetic information in the electronic health record. J Am Med Inform Assoc. 2015;22:1231–1242.

27. Bizon C, Ahalt S, Fecho K, Nassar N, Schmitt CP, Scott E, Wilhelmsen K. Technologies for genomic medicine: CANVAS and AnnoBot, solutions for genomic variant annotation. RENCI Technical Report TR-14-04. RENCI, University of North Carolina at Chapel Hill, 2014. doi: 10.7921/G0QN64N3. http://renci.org/technical-reports/tr-14-04-canvas-and-annobot-solutions-for-genomic-variant-annotation/.

28. Owen P, Ahalt S, Berg J, Coyle J, Evans J, Fecho K, Gillis D, Schmitt CC, Young D, Wilhelmsen K. Technologies for genomic medicine: the GMW, a genetic medical workflow engine. RENCI Technical Report TR-14-02. RENCI, University of North Carolina at Chapel Hill, 2014. doi:10.7921/G0KW5CXC. http://renci.org/technical-reports/tr-14-02-the-gmw-a-genetic-medical-workflow-engine/.

29. Reilly J, Ahalt S, Fecho K, Jones C, McGee J, Roach J, Schmitt CP, Wilhelmsen KC. Technologies for genomic medicine. MaPSeq, a computational and analytical workflow manager for downstream genomic sequencing. RENCI Technical Report TR-14-03. RENCI, University of North Carolina at Chapel Hill, 2014. doi: 10.7921/G0VD6WCF. http://renci.org/technical-reports/mapseq-computational-and-analytical-workflowmanager/.

30. Reilly J, Ahalt S, McGee J, Owen P, Schmitt C, Wilhelmsen K. MaPSeq, a service-oriented architecture for genomics research within an academic biomedical research institution. Informatics 2015;2:20–30. http://www.mdpi.com/2227-9709/2/3/20.

31. Evans JP, Berg JS, Olshan AF, Magnuson T, Rimer BK. We screen newborns, don't we? Realizing the promise of public health genomics. Genet Med. 2013;15(5):332–334.

32. Mane KK, Bizon C, Schmitt C, Owen P, Burchett B, Pietrobon R, Gersing K. VisualDecisionLinc: a visual analytics approach for comparative effectiveness–based clinical decision support in psychiatry. J Biomed Inform. 2012;45(1):101–106.

33. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal Health Records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc. 2006;13(2):121–126.

34. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363(6):501–504.

35. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;339(6117):321–324.

36. de Montjoye YA, De Montjoye YV, Radaelli L, Singh VK, Pentland AS. Unique in the shopping mall: on the reidentifiability of credit card metadata. Science. 2015;347(6221):536–539.

37. Wilhelmsen K, Schmitt CP, Fecho K. Factors influencing data archival of large-scale genomic data sets. RENCI Technical Report TR-13-03. RENCI, University of North Carolina at Chapel Hill, 2013. doi: 10.7921/GOMW2F25. www.renci.org/publications/technical-reports.

38. Rajasekar A, Moore R, Wan M, and Schroeder W. Policy-based distributed data management systems. J Digital Information. 2010;11(1).

39. Rajasekar A, Wan M, Moore R, and Schroeder W. Micro-Services: a service-oriented paradigm for scalable, distributed data management. *Data Intensive Distributed Computing*. Hershey, PA: Information Science Reference; 2012:74-93.

40. Schmitt C, Wilhelmsen K, Krishnamurthy A, Ahalt S, Fecho K. Security and privacy in the era of big Ddta: iRODS, a technological solution to the challenge of implementing security and privacy policies and procedures. RENCI/National Consortium for Data Science white paper. RENCI, University of North Carolina at Chapel Hill, 2013. doi: 10.7921/G0H41PBR. http://www.renci.org/wp-content/uploads/2014/02/0313WhitePaperiRODS.pdf.

41. Fortner B, Ahalt S, Coposky J, Fecho K, Krishnamurthy A, Moore R, Rajasekar A, Schmitt C, Schroeder W. Control Your Data: iRODS, the integrated Rule-Oriented Data System. RENCI, University of North Carolina at Chapel Hill, 2014. Available at: http://renci.org/wp-content/uploads/2014/07/0214WhitePaper-iRODS-2FINAL-v6.pdf.

42. Schmitt C, Shoffner M, Owen P, Wang X, Lamm B, Mostafa J, Barker M, Krishnamurthy A, Wilhelmsen K, Ahalt S, Fecho K. Security and privacy in the era of big data: The SMW, a technological solution to the challenge of data leakage. RENCI/National Consortium for Data Science white paper. RENCI, University of North Carolina at Chapel Hill, 2013. doi: 10.7921/G0WD3XHT. http://www.renci.org/wpcontent/uploads/2014/02/0213WhitePaper-SMW.pdf.

43. Shoffner M, Owen P, Mostafa J, Lamm B, Wang X, Schmitt CP, Ahalt SP. The Secure Medical Research Workspace: An IT infrastructure to enable secure research on clinical data. Clin Transl Sci 2014;6(3), 222–225. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682797.