



Published in final edited form as:

JAMA. 2013 March 27; 309(12): . doi:10.1001/jama.2013.1579.

## Crossing the Omic Chasm: A Time for Omic Ancillary Systems

**Justin Starren, MD, PhD,**

Division of Health and Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

**Marc S. Williams, MD, and**

Genomic Medicine Institute, Geisinger Clinic, Danville, Pennsylvania

**Erwin P. Bottinger, MD**

The Charles Bronfman Institute for Personalized Medicine, Mount Sinai School of Medicine, New York, New York

---

### Viewpoint:

Despite the information gains from Genome-Wide Association Studies (GWAS) and Next-Generation Sequencing (NGS), there remains a chasm between this scientific knowledge and daily clinical practice. Leveraging recent advances in genomics to improve patient care will require Electronic Health Record (EHR) systems that incorporate genomic clinical decision support (CDS). The Electronic Medical Records and Genomics (eMERGE)<sup>1,2</sup> consortium is bridging this chasm by developing interoperable systems that can integrate large-scale genomic data with clinical workflows. We concur with a recent Institute of Medicine report<sup>3</sup>, stating that the current document-centric approach to omic (e.g. genomic, epigenomic, proteomic, metabolomics, and others) data will not scale, and that the storage of raw omic data in current-generation EHRs is not feasible. While commercial EHRs may eventually evolve to handle omic data efficiently, dedicated omic ancillary systems will be essential in the interim.

Historically the term “EHR” refers to the system that is used in day-to-day patient care, while systems that collect and manage specialized data, such as laboratory, pharmacy, and imaging, are considered “ancillary” systems. Before discussing how an omic ancillary would function, it is useful to consider how omic data differs from conventional health data. The simplest difference is the amount of data. Current EHRs are optimized to manage large numbers of discrete and actionable items to facilitate clinical care. They are not designed to store large blocks of data that do not require rapid access. When diagnostic tests create large data sets, ancillary systems are developed to manage the data and only a subset of clinically relevant information is transferred to the EHR. The archetypal example is imaging. Radiologic images are stored in a Picture Archiving and Communication System (PACS). Radiologists interpret the images to generate a report. Only the text report is stored in the EHR’s database. An EHR may interface with a PACS to display images, but does not store them. This dichotomy is evident by comparing the amount of data stored in the EHR versus ancillary systems. At Northwestern University, the EHR averages 375 KB per patient, while the PACS averages 104 MB per patient (a 277-fold difference). Typical NGS identifies 3–10 million variants per individual,<sup>4</sup> requiring 5–10GB of storage (50-fold greater than

---

**Corresponding Author:** Justin Starren, MD, PhD, FACMI, Division of Health and Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, 750 N. Lake Shore Drive, 11th floor, Chicago IL 60611, Phone: (312) 503-2388, Fax: (312) 503-5388, Justin.starren@northwestern.edu.

The authors have no competing interests to declare.

imaging). Compression may reduce the absolute storage requirement, but will not reduce the number of variants to be considered.

One maxim that is taught to medical students is never to order any test unless the results of the test may affect a treatment decision. The implication of this maxim is that clinicians know how they will interpret each piece of information *before* that information is collected. Most clinical laboratory tests measure variable parameters (e.g. serum sodium) compared to set reference values, allowing interpretation using the relevant clinical context. It is unlikely that our understanding of sodium metabolism will change so radically that we would have to reinterpret historical sodium values.

Omic data is different. An individual's germline genetic sequence changes very little over a lifetime but our understanding of that sequence is changing rapidly. For years the DNA between coding regions was called "junk" but we now suspect that this DNA plays an important role in gene regulation.<sup>5,6</sup> The 1000 Genomes project has identified tens of millions of different genomic variants the clinical significance of which is mostly unknown, but our understanding is rapidly changing. Unlike serum sodium, the clinical implications of an NGS obtained today will keep changing for years as knowledge evolves. This necessitates systems that dynamically reanalyze and reinterpret stored static genomic results in the context of evolving knowledge.

Figure 1 shows the data paths that omic data may take from collection to use. Today, most genomic results are delivered as textual reports (Route A). Small panels of genes can be handled as conventional laboratory results (Route B).<sup>7</sup> While some results will likely continue in these modes, NGS will overwhelm both routes. Conventional human-centric approaches to genomic results do not support reinterpretation as new knowledge emerges and will be gradually replaced by computer-centric approaches.

Increasingly NGS and other high-dimensionality omic testing will become routine, creating large, complex data sets, and setting the stage for "omic" ancillary systems. This approach adds value by providing a location to store variants of unknown significance until enough knowledge emerges to move these variants into the clinic. Three paths could facilitate transfer of actionable attributes to the EHR:

- Route C results of the genomic analysis are manually reviewed, converted to a textual report and presented to the clinician.
- Route D creation of "computable observations" that are stored within the EHR where the observations can be used to trigger conventional CDS rules.
- Route E incorporates an external CDS system that is queried by the EHR at appropriate points in the clinical workflow. Although that CDS is shown as part of the omic ancillary it is also possible that this CDS could be external to the organization.

Large organizations will likely operate their own omics ancillary systems, in the same way that they maintain other ancillaries. For smaller practices, reference laboratories may add omic ancillary services to their existing services. The number of clinically significant variants is currently limited, but the availability of affordable NGS will greatly accelerate this flow. Omic ancillaries are one way to bridge the omic chasm without waiting for an entirely new generation of EHRs to emerge.

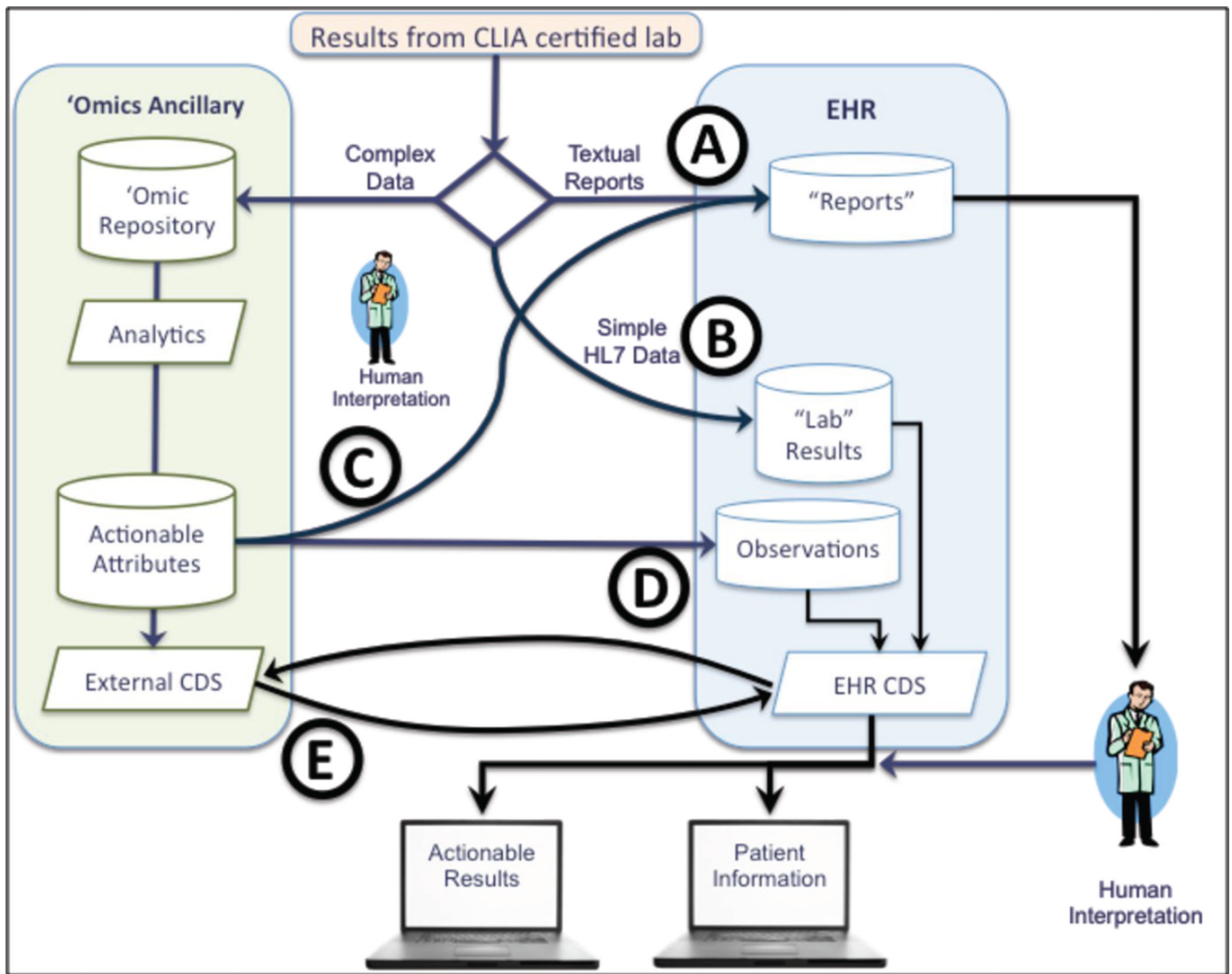
## Acknowledgments

The authors thank all the members of the eMERGE EHR Integration workgroup for input on this article. The eMERGE Network was initiated and funded by NHGRI through the following grants: U01HG006828 (Cincinnati

Children's Hospital Medical Center/Harvard); U01HG006830 (Children's Hospital of Philadelphia); U01HG006389 (Essentia Institute of Rural Health); U01HG006382 (Geisinger Clinic); U01HG006375 (Group Health Cooperative); U01HG06379 (Mayo Clinic.); U01HG006380 (Mount Sinai School of Medicine); U01HG006388 (Northwestern University); U01HG006378 (Vanderbilt University); and U01HG006385 (Vanderbilt University serving as the Coordinating Center).

## References

1. McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011; 4:13. [PubMed: 21269473]
2. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med*. 2011 Apr 20.3(79):79re71.
3. Integrating Large-Scale Genomic Information into Clinical Practice: Workshop Summary. The National Academies Press; 2012. Institute of Medicine Roundtable on Translating Genomic-Based Research for Health.
4. Pelak K, Shianna K, Ge D, Maia J, Zhu M, et al. The Characterization of Twenty Sequenced Human Genomes. *PLoS Genet*. 2010; 6(9):e1001111. [PubMed: 20838461]
5. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6; 489(7414):57–74. [PubMed: 22955616]
6. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490(7418):61–70. [PubMed: 23000897]
7. HL7 Clinical Genomics Committee. *Clinical Genomics*. <http://www.hl7.org/Special/committees/clingenomics/index.cfm>.



**Figure 1.** Integration of Genomic Data with EHR depends on the format and complexity of the data. *Note that data paths convey high-level data flows and do not necessarily imply point-to-point connections.*