J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Viewpoint

A conceptual model for translating omic data into clinical action

Timothy M. Herr[1], Suzette J. Bielinski[2], Erwin Bottinger[3], Ariel Brautbar[4], Murray Brilliant[5], Christopher G. Chute[6], Joshua Denny[7], Robert R. Freimuth[2], Andrea Hartzler[8], Joseph Kannry[9], Isaac S. Kohane[10], Iftikhar J. Kullo[11], Simon Lin[12], Jyotishman Pathak[2], Peggy Peissig[13], Jill Pulley[14], James Ralston[8], Luke Rasmussen[1], Dan Roden[14], Gerard Tromp[15], Marc S. Williams[16], Justin Starren[1]

[1]Department of Preventive Medicine, Division of Health and Biomedical Informatics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, [2]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, [3]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine, Mount Sinai, New York, [4]Division of Genetics and Endocrinology, Cook Children's Medical Center, Fort Worth, Texas, [5]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, [6]Division of General Internal Medicine, Johns Hopkins University, Baltimore, Maryland, [7]Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, [8]Group Health Research Institute, Seattle, Washington, [9]Icahn School of Medicine, Mount Sinai, New York, [10]Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, [11]Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, [12]Nationwide Children's Hospital, Columbus, Ohio, [13]Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, [14]Vanderbilt University School of Medicine, Nashville, Tennessee, [15]Weis Center for Research, [16]Genomic Medicine Institute, Geisinger Health System, Danville, Pennsylvania, USA

E-mail: *Timothy M. Herr - tim.herr@northwestern.edu
*Corresponding author

Received: 31 March 2015 Accepted: 01 July 2015 Published: 31 August 2015

Abstract

Genomic, proteomic, epigenomic, and other "omic" data have the potential to enable precision medicine, also commonly referred to as personalized medicine. The volume and complexity of omic data are rapidly overwhelming human cognitive capacity, requiring innovative approaches to translate such data into patient care. Here, we outline a conceptual model for the application of omic data in the clinical context, called "the omic funnel." This model parallels the classic "Data, Information, Knowledge, Wisdom pyramid" and adds context for how to move between each successive layer. Its goal is to allow informaticians, researchers, and clinicians to approach the problem of translating omic data from bench to bedside, by using discrete steps with clearly defined needs. Such an approach can facilitate the development of modular and interoperable software that can bring precision medicine into widespread practice.

Key words: Genomic medicine, personalized health care, precision medicine

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.163985

Quick Response Code:

INTRODUCTION

A wide variety of high-throughput technologies is becoming available for clinical diagnosis and care. These include various genomic technologies such as microarrays, targeted gene capture chips, whole exome sequencing, and whole genome sequencing. Other data sources, such as epigenomic, proteomic, metabolomic, and microbiomic, are also becoming available. The sheer volume of data involved in omic analyses[1] and difficulty of interpretation[2] makes it challenging for clinicians to obtain and apply related knowledge. These data types overwhelm any individual clinician's cognitive capacity.[3] Nonetheless, the use of these new data types in the clinical setting could provide

This article may be cited as:
Herr TM, Bielinski SJ, Bottinger E, Brautbar A, Brilliant M, Chute CG, *et al*. A conceptual model for translating omic data into clinical action. J Pathol Inform 2015;6:46.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2015/6/1/46/163985

valuable insight and rapidly advance the practice of precision medicine. Although integrating each of these data types in an electronic health record (EHR) presents novel challenges, the commonalities allow us to refer to these data collectively as "omic." Here, we present a conceptual model for omic data management in the clinical context. This conceptual model serves to inform and complement our implementation-based model.[4] It is also consistent with the National Human Genome Research Institute's (NHGRI) "base pairs to bedside" vision for genomic medicine[5] and the White House's recently announced Precision Medicine Initiative.[6]

This model is based on a series of discussions from within the EHR integration workgroup of the Electronic Medical Records and Genomics (eMERGE) Network,[7] a consortium funded by the NHGRI to study the use of genomic data in research and health care. We have observed that both clinicians and current-generation EHRs struggle with the volume of data produced by omic analyses.[8] Therefore, the challenge is to reduce omic data that may contain billions of individual values into a small number of clinically actionable recommendations.

The eMERGE discussions led to two main insights. The first was that we could learn from other data-intensive clinical information sources like radiology. These sources frequently employ ancillary systems to manage large volumes of data, implying that an "omic ancillary" system would likely be needed.[4] The second insight was that conversion from raw data to actionable knowledge would require multiple external knowledge sources. Envisioning these multiple external sources as sequential filters resulted in the concept of an "omic funnel" [Figure 1].

THE OMIC FUNNEL

The omic funnel aligns with the classic "Data, Information, Knowledge, Wisdom (DIKW) pyramid" from information science[9] [Figure 2]. The DIKW pyramid is a hierarchy progressing from Data to Information, Knowledge, and Wisdom. The progression up each step of the pyramid is based on the addition of context to allow interpretation. In other words, data in context becomes information. Information in context becomes knowledge. Similarly, omic data is successively refined through the application of context.

The DIKW pyramid has previously been applied to other subdomains of the biomedical field. For instance, it has influenced machine learning researchers working with patient databases in an effort to discover new knowledge from large quantities of clinical data.[10] Here, we adapt the same framework to distill clinical knowledge from large volumes of omic data. The traditional DIKW layers are represented in our conceptual model but are now specific to omics. The "omic data" layer represents data of various forms, including output from high-throughput sequencing platforms, methylation data, or tissue arrays. In the example of genomic data, this may be a sequence of letters representing an individual's entire genome, contained in a text file. Because this layer represents an overwhelming amount of data to expect anyone to act upon, it must be filtered and processed as it moves to the subsequent layers of the funnel.

The "biological information" layer contains information about the biological state of individuals. This information can take many forms, such as single-nucleotide polymorphisms (SNPs), gene expression levels, or copy number variations. In genomics, biological information could be represented within a variant call format file, which, instead of carrying the entire genome, contains information about where the individual's genome varies from a reference sequence. Though many have predicted effects, the majority of variants currently have no validated clinical significance.[11,12] Information with unknown or uncertain significance is rarely helpful in the clinical setting, so such information must be filtered for actionability.

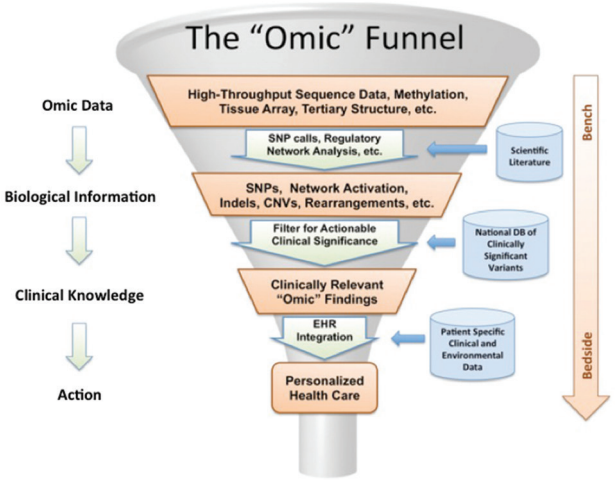

Figure 1: The omic funnel

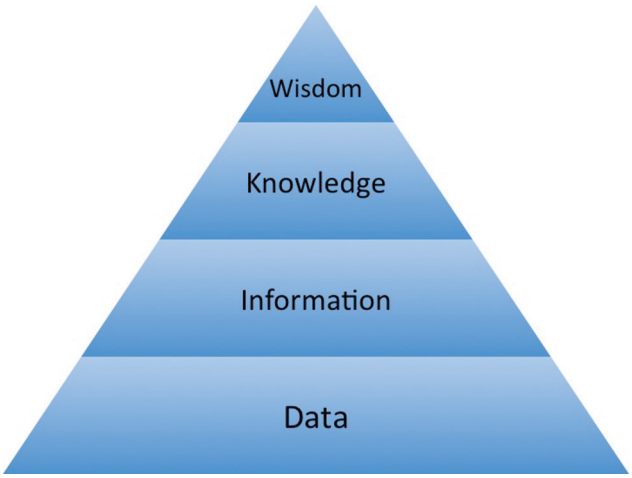

Figure 2: The data, information, knowledge, and wisdom pyramid

The "clinical knowledge" layer represents knowledge that is relevant to the clinical setting in that it can be acted upon during patient care. In other words, this knowledge will include clinically relevant omic associations. Such knowledge can be represented in a variety of formats, such as a textual report or discrete data elements entered into the EHR through HL7. In genomics, clinical knowledge could be a *CYP2C9* or *TPMT* genotype, which has known pharmacogenomic (PGx) associations, in combination with a clinical recommendation. In some cases, this knowledge may be actionable on its own, but in other cases, it may need to be combined with additional clinical data to be truly applicable.

Finally, the "action" layer represents methods by which clinical knowledge is translated to the bedside and applied to change clinician behavior. Clinical knowledge derived from omic data will be considered in treatment and combined with other clinical factors to personalize care. In genomics, action could refer to the use of a patient's *CYP2C9* and *VKORC1* status, along with age, weight, smoking status, and other clinical indicators, to individualize the dose of a new warfarin prescription.[13]

This layered approach is analogous to the open systems interconnection model (OSI).[14] The OSI model allows for modularized computing by defining distinct architectural layers. These layers each exist independently with localized functions and communicate with each other through defined protocols. Similarly, the omic funnel allows data, information, knowledge, and action to exist independently. The modularized software could then be designed for each layer and the transitions between them.

FROM DATA TO ACTION

Transitioning between the layers of the omic funnel model is difficult in practice and requires collaboration between multiple parties. One cannot spontaneously jump from a complete genetic sequence to a list of SNPs. Nor can one view a list of SNPs and instantly recognize clinically relevant genotypes. It is also unrealistic to provide raw omic data to clinicians and expect them to be able to act on this data in practice. Instead, it is necessary to have an infrastructure in place to support each of the necessary transitions from bench to bedside.

The first transition is from raw omic data to biological information. This requires basic research into the nature of individual genes, proteins, and epigenetic features. These results are then vetted and published in the scientific literature to form a basis for clinical investigation. In the past, gene discovery and analysis were largely performed through targeted candidate gene studies. Today, with the advent of high-throughput sequencing, whole-genome variant analyses, genome-wide association studies, and RNA-seq analyses are commonly used. In the future, research that goes beyond the genome by including epigenomic, proteomic, and other datasets, will become more prevalent.

The next transition is from biological information to clinically relevant knowledge. The number of clinically significant variants is currently small,[15] but this number will continue to grow. Moreover, our understanding of the functional effect of variants will change over time, and clinical recommendations will be updated accordingly. In the past, omically-driven knowledge was rarely used in the clinical setting, so it was unnecessary to centrally catalog. Today, genome-driven care is beginning to take hold in areas such as PGx-based drug prescribing. This places a significant burden on provider organizations to maintain current, accurate knowledge of the field. We believe that it will be impossible for any single provider organization to catalog all relevant variants and keep them up to date. Instead, outside organizations will be needed to help track the expanding knowledge base. To this end, the Clinical Pharmacogenetics Implementation Consortium (CPIC) is developing a central repository of evidence-based guidelines for clinically actionable gene-drug interactions.[16] The NHGRI also awarded over $25 million in grants in 2013 for ClinGen, an effort to create a central repository containing clinically relevant genetic variants.[17] Such efforts have the potential to remove the burden of maintaining clinically relevant, omic-derived knowledge from individual providers.

The final transition is to turn clinical knowledge into action. This can often be a complicated process incorporating multiple data points. Take, for example, the algorithms that are currently available for warfarin dosing.[13] In the past, clinicians had to manually run the algorithms and calculate dosages by hand (an onerous and error-prone process). This represents an ideal application for clinical decision support (CDS) integrated into the EHR. Today, tools such as WarfarinDosing.org have automated the calculation process and organizations like eMERGE have begun to implement PGx-driven CDS tools in clinical workflows on a limited basis.[18,19] In the future, CDS will be a powerful tool when it is driven by both local clinical data and easily accessible knowledge from databases like those being created by CPIC and ClinGen.

When achieved, such carefully designed CDS presented at the time of clinical action is the critical component that will reduce the cognitive overload clinicians would otherwise experience when presented with omic data. However, making the transition from clinical knowledge to action through CDS will require a computable knowledge format. Similar work has been done with drug-drug interaction knowledge.

With the SFINX database, drug-drug interaction knowledge is coded and stored in a format that can be shared and integrated into CDS systems.[20] This approach could serve as a model for omic knowledge. For example, the CPIC guideline for clopidogrel dosing breaks therapeutic recommendations down by poor, intermediate, extensive, or ultrarapid metabolizer status, determined by genotype. Each status has a recommendation such as "alternative antiplatelet therapy (if no contraindication); e.g. prasugrel, ticagrelor."[21] However, this knowledge is currently only available through journal publications or in a web format on PharmGKB.org. If this were available in a standard format that CDS systems can recognize, then it could be directly integrated into clinician workflows with significantly less effort.

CONCLUSION

Whereas previous literature generally focused on practical considerations for individual steps in the translation of omic data to patient care,[4,8,22] the model presented here serves as a generalized conceptual framework in which to understand the end-to-end translation of omic data from bench to bedside. There are likely to be many different software and data management architectures and strategies employed to implement these transitions in practice. Even so, our conceptual model provides a step-by-step process to filter an overwhelming amount of complex omic data down to clinical action. We believe that explicitly acknowledging these different transitions will aid the creation of modular, interoperable software solutions. The difficult work of creating practical, real world standards and tools for the transition between each of the layers is early, but under way. Continued biological research, comprehensive electronic knowledge bases, and robust CDS tools are all necessary to translate bench data to the bedside.

Financial Support and Sponsorship

The eMERGE Network was initiated and funded by NHGRI through the following grants: U01HG006828 (Cincinnati Children's Hospital Medical Center/Boston Children's Hospital); U01HG006830 (Children's Hospital of Philadelphia); U01HG006389 (Essentia Institute of Rural Health, Marshfield Clinic Research Foundation and Pennsylvania State University); U01HG006382 (Geisinger Clinic); U01HG006375 (Group Health Cooperative/University of Washington); U01HG006379 (Mayo Clinic); U01HG006380 (Icahn School of Medicine at Mount Sinai); U01HG006388 (Northwestern University); U01HG006378 (Vanderbilt University Medical Center); and U01HG006385 (Vanderbilt University Medical Center serving as the Coordinating Center).

Conflicts of Interest

There are no conflicts of interest.

REFERENCES

1. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren J, Hakonarson H, *et al.* Practical challenges in integrating genomic data into the electronic health record. Genet Med 2013;15:772-8.
2. 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.
3. Stead WW, Searle JR, Fessler HE, Smith JW, Shortliffe EH. Biomedical informatics: Changing what physicians need to know and how they learn. Acad Med 2011;86:429-34.
4. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: A time for omic ancillary systems. JAMA 2013;309:1237-8.
5. Green ED, Guyer MS; National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. Nature 2011;470:204-13.
6. The White House. FACT SHEET: President Obama's Precision Medicine Initiative. Available from: http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative. [Last updated on 2015 Jan 30; Last cited on 2015 Jan 30].
7. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, *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.
8. 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;45:419-22.
9. Rowley J. The wisdom hierarchy: Representations of the DIKW hierarchy. J Inf Sci 2007;33:163-80.
10. Meyfroidt G, Güiza F, Ramon J, Bruynooghe M. Machine learning techniques to examine large patient databases. Best Pract Res Clin Anaesthesiol 2009;23:127-43.
11. Khoury MJ, Coates RJ, Evans JP. Evidence-based classification of recommendations on use of genomic tests in clinical practice: Dealing with insufficient evidence. Genet Med 2010;12:680-3.
12. 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;13:499-504.
13. International Warfarin Pharmacogenetics Consortium, Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, *et al.* Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 2009;360:753-64.
14. Zimmermann H. OSI reference model – The ISO model of architecture for open systems interconnection. IEEE Trans Commun 1980;28:425-32.
15. Amstutz U, Carleton BC. Pharmacogenetic testing: Time for clinical practice guidelines. Clin Pharmacol Ther 2011;89:924-7.
16. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the pharmacogenomics research network. Clin Pharmacol Ther 2011;89:464-7.
17. National Institutes of Health. New NIH-funded Resource Focuses on Use of Genomic Variants in Medical Care. Available from: http://www.nih.gov/news/health/sep2013/nhgri-25.htm. [Last updated on 2013 Sep 25; Last cited on 2015 Jan 27].
18. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future. Genet Med 2013;15:761-71.
19. Rasmussen-Torvik LJ, Stallings SC, Gordon AS, Almoguera B, Basford MA, Bielinski SJ, *et al.* Design and anticipated outcomes of the eMERGE-PGx project: A multicenter pilot for preemptive pharmacogenomics in electronic health record systems. Clin Pharmacol Ther 2014;96:482-9.
20. Böttiger Y, Laine K, Andersson ML, Korhonen T, Molin B, Ovesjö ML, *et al.* SFINX-a drug-drug interaction database designed for clinical decision support systems. Eur J Clin Pharmacol 2009;65:627-33.
21. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. Clin Pharmacol Ther 2013;94:317-23.
22. Welch BM, Eilbeck K, Del Fiol G, Meyer LJ, Kawamoto K. Technical desiderata for the integration of genomic data with clinical decision support. J Biomed Inform 2014;51:3-7.