

Variant Validity (Selected vs. General Population)

Paul R. Billings, Nalini Raghavachari, and Geetha Senthil*

January 22, 2015

**The authors are participants in the activities of the IOM Roundtable on Translating Genomic-Based Research for Health.*

The views expressed in this discussion paper are those of the authors and not necessarily of the authors' organizations or of the Institute of Medicine. The paper is intended to help inform and stimulate discussion. It has not been subjected to the review procedures of the Institute of Medicine and is not a report of the Institute of Medicine or of the National Research Council.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the nation • Improving health

Copyright 2015 by the National Academy of Sciences. All rights reserved.

The types of evidence needed to support the use of genome sequencing in the clinic varies by stakeholder and circumstance. In this IOM series, seven individually authored commentaries explore this important issue, discussing the challenges involved in and opportunities for moving clinical sequencing forward appropriately and effectively.

Variant Validity (Selected vs. General Population)

Paul R. Billings, Thermo Fisher Scientific; Nalini Raghavachari, National Institute on Aging; and Geetha Senthil, National Institute of Mental Health^{1,2}

Hippocrates, the father of medicine, advocated “Declare the past, diagnose the present, and foretell the future” as a good clinical practice in medicine (Hippocrates, 1993). Translating this quote, written 2,500 years ago, in the current context of advanced genomic tools and technologies would emphasize the importance of sequencing the genome, identifying and clinically annotating disease-associated variants, and foretelling the future with genomics providing tailored individualized therapy.

Advances in genome sequencing represent an unparalleled opportunity to examine the genomic landscape of an individual and potentially identify genetic variants that are relevant for the diagnosis and treatment of disease. Among the available genomic tools, unbiased next-generation sequencing (NGS) methods such as exome and genome sequencing and targeted sequencing are revolutionizing the study of genetic variation, providing a better understanding of the role that genes play in disease development and progression for research and clinical practice (Bennett and Farah, 2014; Biswas et al., 2014; Brion et al., 2014; Dello Russo et al., 2014; Javitt and Carner, 2014; Kamalakaran et al., 2013). In addition to NGS methods, nonsequencing array-based molecular methods offer a cost-efficient platform for screening and defining genetic variation in an efficient manner (Katsanis and Katsanis, 2013).

Once variants are noted in a human genome, they need to be interpreted so that they can be properly applied in clinical practice. The clinical interpretation of data should fall into any one of the following categories (Katsanis and Katsanis, 2013; Kearney et al., 2011; Richards et al., 2008; Zhang and Wang, 2012):

1. Disease causing: a variation in sequence has been reported previously and is recognized as being causative of the disorder;
2. Likely disease causing: a variation in sequence has not been reported previously but is of a particular type that is expected to be causative of the disorder;
3. Possibly disease causing: a variation in sequence has not been reported previously but is of a particular type that may or may not be causative of the disorder;

¹ The authors are participants in the activities of the IOM Roundtable on Translating Genomic-Based Research for Health.

² Suggested citation: Billings, P. R., N. Raghavachari, and G. Senthil. 2015. *Variant validity (selected vs. general population)*. Discussion Paper, Institute of Medicine, Washington, DC. <http://nam.edu/wp-content/uploads/2015/06/VariantValidity.pdf>

4. Likely not disease causing: a variation in sequence has not been reported previously and is likely not causative of the disorder;
5. Not disease causing: a variation in sequence has been reported previously and is recognized as not being causative of the disorder;
6. Protective variant: sequence variation has previously been reported to be associated with disease-free phenotype or lesser susceptibility to diseases or pleiotropic effect of the variant. Such protective variants are especially of interest in diagnosis and treatment because of their potential predictive/prognostic value for co-morbid or multiple chronic conditions (Bougle et al., 2012; Guella et al., 2010; Saade et al., 2011; Scartezini et al., 2007; Stitzel et al., 2014);
7. Variant of unknown clinical significance: a variation in sequence is not known or expected to be causative of the disorder but is identified as being associated with a clinical presentation.

High-quality clinical validation of NGS results, expressed as their clinical validity, is a crucial component of their inculcation into medical practice. Clinical validity is the accuracy of the data resulting from a test in identifying, measuring, or predicting the presence or absence of a clinical condition (with potential treatment implications) or predisposition (Alberg et al., 2004; Burke, 2009; Burke et al., 2002; Burke et al., 2007). Typical formulations of clinical validity include a test's sensitivity and specificity; its positive or negative predictive value; determinations of an assay's Area under the Receiver Operating Characteristic Curve—AUROC (Alberg et al., 2004; Burke et al., 2002; Burke et al., 2007); and other measures. Clinically valid tests impact risk assessment, differential diagnoses, and many other aspects of an increasingly precise delivery of medical care.

Historically, particularly for biochemical disorders, correlation of genetic analysis with established familial findings and known protein changes allowed rapid clinical validation of DNA tests. For variants related to sickle cell disease or phenylketonuria, newly identified genetic variants could be assessed with orthogonal assays and then the associated single-nucleotide polymorphisms (SNPs) could be routinely ascertained by genotyping or, when developed, sequencing. Once Sanger or NGS methods became feasible for known disease-causing loci, newly identified variants could be evaluated by their impact on the cornerstone biochemical activity and/or clinical phenotype. With time, our ability to identify potentially clinically relevant SNPs has improved, while other measurements of proven disease causation remain cumbersome. A general process for clinical validation that depends solely on genetic data has evolved using comparative studies on new variants to SNP databases. The first step in this type of clinical validation was usually a Genome Wide Association Study or family study that focused on a variant(s) and associated it with disease. Then, a confirmatory experiment in a similar population or type of family would occur. Finally, the ability of the genetic variant to correlate with disease in other subpopulations or a “general,” all-comers approach was assessed.

In current clinical practice, associating the identified and interpreted variants as causal for diseases and relevant to treatment choice can require extensive evaluation to determine precisely the clinical validity of the test results in often highly mixed human populations. Although this is a major challenge in terms of experimental design, time, and cost, a test that leads to a diagnosis and follow-up treatment may have tremendous utility from both a patient's and clinician's perspective. How can we overcome this challenge?

1. Several outcomes of current protocols for clinical validation are possible: similar disease, different disease, no disease, or uncertain. But once a variant is well established as causal in even a small cohort, thereby implicating a biological pathway(s), the probability that the same genetic variant could be causative for disease in other groups that make up the general population is much higher. Thus the burden of proof for validating a significant variant classification in other populations is *not* de novo but rather modified by the earlier results.
2. In an effort to confirm clinically relevant variants, we need to replicate and validate the findings in a stratified population cohort to rule out confounding factors associated with the disease.
3. As further confirmation, we could replicate findings in a general population and validate the role of the variant in disease pathogenesis. But a lower evidentiary standard should be applied that considers variable phenotypes, allelic frequencies, genomic redundancies, confounding environmental influencers, and so forth.
4. In addition, a different technology or platform could be applied to cross-validate the identified variant in association to the disease.

The application of these and other solutions could speed the translation of clinically valid test results, demonstrated in select groups, to other groups that make up the general population. For those in which the variant is pathological, this more rapid validation could have great significance.

REFERENCES

- Alberg, A. J., J. W. Park, B. W. Hager, M. V. Brock, and M. Diener-West. 2004. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine* 19(5 Pt 1):460–465.
- Bennett, N. C., and C. S. Farah. 2014. Next-generation sequencing in clinical oncology: Next steps towards clinical validation. *Cancers (Basel)* 6(4):2296–2312.
- Biswas, A., V. R. Rao, S. Seth, and S. K. Maulik. 2014. Next generation sequencing in cardiomyopathy: Towards personalized genomics and medicine. *Molecular Biology Reports* 41(8):4881–4888.
- Bougle, A., A. Max, N. Mongardon, D. Grimaldi, F. Pene, C. Rousseau, J. D. Chiche, J. P. Bedos, E. Vicaut, and J. P. Mira. 2012. Protective effects of FCGR2A polymorphism in invasive pneumococcal diseases. *Chest* 142(6):1474–1481.
- Brion, M., A. Blanco-Verea, B. Sobrino, M. Santori, R. Gil, E. Ramos-Luis, M. Martinez, J. Amigo, and A. Carracedo. 2014. Next generation sequencing challenges in the analysis of cardiac sudden death due to arrhythmogenic disorders. *Electrophoresis* 35(21–22):3111–3116.
- Burke, W. 2009. Clinical validity and clinical utility of genetic tests. *Current Protocols in Human Genetics* 60:9.15.1–9.15.7
- Burke, W., D. Atkins, M. Gwinn, A. Guttmacher, J. Haddow, J. Lau, G. Palomaki, N. Press, C. S. Richards, L. Wideroff, and G. L. Wiesner. 2002. Genetic test evaluation: Information

- needs of clinicians, policy makers, and the public. *American Journal of Epidemiology* 156(4):311–318.
- Burke, W., R. L. Zimmern, and M. Kroese. 2007. Defining purpose: A key step in genetic test evaluation. *Genetics in Medicine* 9(10):675–681.
- Dello Russo, C., G. Di Giacomo, A. Mesoraca, L. D'Emidio, P. Iaconianni, E. Minutolo, A. Lippa, and C. Giorlandino. 2014. Next generation sequencing in the identification of a rare genetic disease from preconceptional couple screening to preimplantation genetic diagnosis. *Journal of Prenatal Medicine* 8(1–2):17–24.
- Guella, I., R. Asselta, D. Ardissino, P. A. Merlini, F. Peyvandi, S. Kathiresan, P. M. Mannucci, M. Tubaro, and S. Duga. 2010. Effects of PCSK9 genetic variants on plasma LDL cholesterol levels and risk of premature myocardial infarction in the Italian population. *Journal of Lipid Reseach* 51(11):3342–3349.
- Hippocrates, C. 1993. [The sources of medical ethics: Hippocrates and his disciples]. *Revue de l'infirmière* 43(8):12–14.
- Javitt, G. H., and K. S. Carner. 2014. Regulation of next generation sequencing. *Journal of Law, Medicine, and Ethics* 42 (Suppl 1):9–21.
- Kamalakaran, S., V. Varadan, A. Janevski, N. Banerjee, D. Tuck, W. R. McCombie, N. Dimitrova, and L. N. Harris. 2013. Translating next generation sequencing to practice: Opportunities and necessary steps. *Molecular Oncology* 7(4):743–755.
- Katsanis, S. H., and N. Katsanis. 2013. Molecular genetic testing and the future of clinical genomics. *Nature Reviews Genetics* 14(6):415–426.
- Kearney, H. M., E. C. Thorland, K. K. Brown, F. Quintero-Rivera, S. T. South, and the Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. 2011. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genetics in Medicine* 13(7):680–685.
- Richards, C. S., S. Bale, D. B. Bellissimo, S. Das, W. W. Grody, M. R. Hegde, E. Lyon, and B. E. Ward. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in Medicine* 10(4):294–300.
- Saade, S., J. B. Cazier, M. Ghassibe-Sabbagh, S. Youhanna, D. A. Badro, Y. Kamatani, J. Hager, J. S. Yeretziyan, G. El-Khazen, M. Haber, A. K. Salloum, B. Douaihy, R. Othman, N. Shasha, S. Kabbani, H. E. Bayeh, E. Chammas, M. Farrall, D. Gauguier, D. E. Platt, and P. A. Zalloua. 2011. Large scale association analysis identifies three susceptibility loci for coronary artery disease. *Public Library of Science One* 6(12):e29427.
- Scartezini, M., C. Hubbart, R. A. Whittall, J. A. Cooper, A. H. Neil, and S. E. Humphries. 2007. The PCSK9 gene R461 variant is associated with lower plasma lipid levels and cardiovascular risk in healthy U.K. men. *Clinical Science (London)* 113(11):435–441.
- Stitzel, N. O., H. H. Won, A. C. Morrison, G. M. Peloso, R. Do, L. A. Lange, P. Fontanillas, N. Gupta, S. Duga, A. Goel, M. Farrall, D. Saleheen, P. Ferrario, I. Konig, R. Asselta, P. A. Merlini, N. Marziliano, M. F. Notarangelo, U. Schick, P. Auer, T. L. Assimes, M. Reilly, R. Wilensky, D. J. Rader, G. K. Hovingh, T. Meitinger, T. Kessler, A. Kastrati, K. L. Laugwitz, D. Siscovick, J. I. Rotter, S. L. Hazen, R. Tracy, S. Cresci, J. Spertus, R. Jackson, S. M. Schwartz, P. Natarajan, J. Crosby, D. Muzny, C. Ballantyne, S. S. Rich, C. J. O'Donnell, G. Abecasis, S. Sunyaev, D. A. Nickerson, J. E. Buring, P. M. Ridker, D. I. Chasman, E. Austin, Z. Ye, I. J. Kullo, P. E. Weeke, C. M. Shaffer, L. A. Bastarache, J. C. Denny, D. M. Roden, C. Palmer, P. Deloukas, D. Y. Lin, Z. Z. Tang, J. Erdmann, H. Schunkert, J. Danesh, J.

Marrugat, R. Elosua, D. Ardissino, R. McPherson, H. Watkins, A. P. Reiner, J. G. Wilson, D. Altshuler, R. A. Gibbs, E. S. Lander, E. Boerwinkle, S. Gabriel, and S. Kathiresan. 2014. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *New England Journal of Medicine* 371(22):2072–2082.

Zhang, V. W., and J. Wang. 2012. Determination of the clinical significance of an unclassified variant. *Methods in Molecular Biology* 837:337–348.