Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics

ACMG Board of Directors[1]

Disclaimer: These recommendations are designed primarily as an educational resource for medical geneticists and other health-care providers, to help them provide quality medical genetic services. Adherence to these recommendations does not necessarily assure a successful medical outcome. These recommendations should not be considered inclusive of all proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, the geneticist should apply his or her own professional judgment to the specific clinical circumstances presented by the individual patient or specimen. It may be prudent, however, to document in the patient's record the rationale for any significant deviation from these recommendations.

There exist 5,000–7,000 rare genetic diseases, each of which harbors considerable clinical variability. None are common individually. In addition, more common diseases with genetic influences may have rare variants associated with them. Vast allelic heterogeneity lies at the foundation of most genetic diseases, the effects of which are compounded by background genomic variation that may further affect clinical presentation.

The considerable variation in clinical presentation and molecular etiology of genetic disorders, coupled with their relative individual rarity, makes it clear that no single provider, laboratory, medical center, state, or even individual country will typically possess sufficient knowledge to deliver the best care for patients in need of care. Even in the relatively rare situation in which pathogenic variants are few (e.g., sickle cell anemia), variants in other alleles may contribute to the genomic variation and clinical manifestations of disease. For more genetically complex conditions such as cystic fibrosis, in spite of decades of study, as many as 10% of cases have a *CFTR* variant so rare that it is represented in only one or two people in current databases, a situation paralleled in many genetic diseases.[1,2]

To ensure that our patients receive the most informed care possible, the American College of Medical Genetics and Genomics advocates for extensive sharing of laboratory and clinical data from individuals who have undergone genomic testing. Information that underpins health-care service delivery should be treated neither as intellectual property nor as a trade secret when other patients may benefit from the knowledge being widely available. It is similarly important for understanding the risks associated with genetic test results that place asymptomatic/presymptomatic individuals at high risk of developing a genetic disease. Sharing data in this precompetitive space will provide both a resource for clinical laboratories interpreting test results and clinical validity data that can benefit device manufacturers developing new tests and testing platforms. Contributing to public clinical databases in the precompetitive space recognizes that information about genetic diseases is dense and accumulating rapidly, and that information science is empowering the use of "big data." Further, the shift to public databases being populated by de-identified case-level information from electronic health records will speed the time to "publication" of what are essentially case reports in real time. This process can also reduce the time period during which one might be able to protect trade secrets. Recognizing the importance of data sharing for both research and clinical care, the National Institutes of Health has established a genomic data-sharing policy for its funded investigators.[3]

Responsible sharing of genomic variant and phenotype data will provide the robust information necessary to improve clinical care and empower device and drug manufacturers that are developing tests and treatments for patients.

- Broad data sharing is necessary and will improve care by making available the best data possible by which:
  - Key clinical attributes of the phenotype of those with genetic diseases can be described
  - The qualitative strength of the association between genetic diseases and the underlying causative genes can be established

[1]American College of Medical Genetics and Genomics, Bethesda, Maryland, USA. Correspondence: Michael S. Watson (mwatson@acmg.net)
The Board of Directors of the American College of Medical Genetics and Genomics approved this statement on 26 September 2016.

- The classification of genomic variants across the range of benign to pathogenic can be established
  - Differences in variant interpretation among laboratories can be reconciled
  - The appropriate classification of variants of uncertain significance can be made
  - Standards used in variant classification can be improved
- Data sharing will provide the scientific community, health-care providers, and industry with the best data on which:
  - Web-based systems for integrated clinical decision support are based
  - Secondary studies using these data are powered
- Data sharing will offer significant financial benefits by which:
  - More standardized approaches to coverage and reimbursement policies can be made
  - The expensive duplication of previously resolved, but unpublished, research efforts currently occurring among pharmaceutical companies can be reduced

The analytical challenges of migrating and integrating clinical and laboratory data across the genome are daunting. Standardization of laboratory and clinical information will enable:

- Data compatibility
- Interoperability between information systems

Importantly, broad data sharing is compatible with the critical imperative of protecting the privacy of individual health-care information and the security of data systems holding that information. For data to be shared safely for patients and providers, systems are required that:

- Ensure the security of databases, whether centralized or federated
- Ensure the privacy of patient and family medical information
- Provide transparency in the documentation of data-sharing transactions

Clinical-grade standards by which claims about gene/disease associations and the clinical significance of variants are made (e.g., data provenance, database versioning, and expert information curation) are central to a shared genomics data system. However, the need to deliver safe and effective care for those with or at risk for rare diseases, despite weak data for most variants and inevitable conflicts in data interpretation, requires balancing regulatory oversight with the need to provide services regardless of how well a rare disease is understood.

Due to the vast amount of data now being generated by genomic testing, genetic diseases will offer the opportunity to develop the framework for a national learning health-care system because the shared experiences of those caring for these patients continually contribute to improvements in delivering services to this population. A learning health-care system that facilitates access to diagnostic, treatment, and outcomes data to inform the care of today's patients requires a paradigm shift in how we share data to be used in research and clinical practice. Academic medical centers have already begun to address how providers within their systems can use information about their patients to benefit other patients. This approach could be made national in scope to the benefit of patients everywhere. The National Institutes of Health has already made such data sharing a priority in the research that it funds. However, to accomplish these goals, and to ensure that the tremendous amounts of information now being generated are not wasted, our community must both demonstrate the will to share data broadly and develop the mechanisms to do so easily. These efforts will require support and participation from clinical laboratories, clinicians, regulatory agencies, researchers, and patients to ensure success in improving patient care through genomic medicine.

DISCLOSURE
The authors declare no conflict of interest.

REFERENCES
1. *The Clinical and Functional Translation of CFTR (CFTR2)*. http://cftr2.org. Accessed 8 November 2016.
2. Rehm HL, Berg JS, Brooks LD, et al.; ClinGen. ClinGen–the Clinical Genome Resource. *N Engl J Med* 2015;372:2235–2242.
3. National Institutes of Health. *NIH Genomic Data Sharing Policy*. http://gds.nih.gov/03policy2.html. Accessed 8 November 2016.