Bedside Back to Bench: Building Bridges between Basic and Clinical Genomic Research

Teri A. Manolio,[1,*] Douglas M. Fowler,[2] Lea M. Starita,[2] Melissa A. Haendel,[3] Daniel G. MacArthur,[4,5] Leslie G. Biesecker,[1] Elizabeth Worthey,[6] Rex L. Chisholm,[7] Eric D. Green,[1] Howard J. Jacob,[6] Howard L. McLeod,[8] Dan Roden,[9] Laura Lyman Rodriguez,[1] Marc S. Williams,[10] Gregory M. Cooper,[6] Nancy J. Cox,[11] Gail E. Herman,[12] Stephen Kingsmore,[13] Cecilia Lo,[14] Cathleen Lutz,[15] Calum A. MacRae,[16] Robert L. Nussbaum,[17] Jose M. Ordovas,[18] Erin M. Ramos,[1] Peter N. Robinson,[19] Wendy S. Rubinstein,[20] Christine Seidman,[21,22,23] Barbara E. Stranger,[24] Haoyi Wang,[25] Monte Westerfield,[26] and Carol Bult[25]

[1]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[3]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239, USA
[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA
[6]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
[7]Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[8]DeBartolo Family Personalized Medicine Institute, Moffitt Cancer Center, Tampa, FL 33612, USA
[9]Department of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA
[10]Genomic Medicine Institute, Geisinger Health System, Danville, PA 17822, USA
[11]Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37203, USA
[12]Institute for Genomic Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA
[13]Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA
[14]Department of Developmental Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 1526, USA
[15]Rare and Orphan Disease Center, Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME 04609, USA
[16]Divisions of Cardiovascular Medicine, Network Medicine and Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA
[17]Invitae Genetics Information and Testing Company, San Francisco, CA 94107, USA
[18]JM-USDA-Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, USA
[19]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA
[20]National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20892, USA
[21]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[22]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[23]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
[24]Section of Genetic Medicine, Department of Medicine, Institute for Genomics and Systems Biology, Center for Data Intensive Science, University of Chicago, Chicago, IL 60637, USA
[25]The Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME 04609, USA
[26]Department of Biology, University of Oregon, Portland, OR 97403, USA
*Correspondence: manolio@nih.gov
http://dx.doi.org/10.1016/j.cell.2017.03.005

SUMMARY

Genome sequencing has revolutionized the diagnosis of genetic diseases. Close collaborations between basic scientists and clinical genomicists are now needed to link genetic variants with disease causation. To facilitate such collaborations, we recommend prioritizing clinically relevant genes for functional studies, developing reference variant-phenotype databases, adopting phenotype description standards, and promoting data sharing.

INTRODUCTION

The use of genome sequencing to diagnose human disease has grown dramatically since the first such reports in 2010 (Cooper, 2015), leading to a surge in reports of novel genetic variants associated with human disease. The accelerating pace of the discovery of such variants, however, is vastly outstripping our ability to understand their potential impact on biologic function and relevance for pathological processes, and an uncomfortably large proportion of variants that are being newly identified is relegated to the indeterminate category of variants of uncertain significance (VUSs) (Cooper, 2015). Advances in technologies for the functional characterization of genetic variants, as well as emerging informatics resources, are presenting new opportunities for basic and clinical scientists to build upon each other's work in a virtuous cycle of bench to bedside and back again. This led the National Human Genome Research Institute, in April 2016, to convene representatives of these communities to explore ways to facilitate basic-clinical collaborations for interpreting VUSs and translate that knowledge

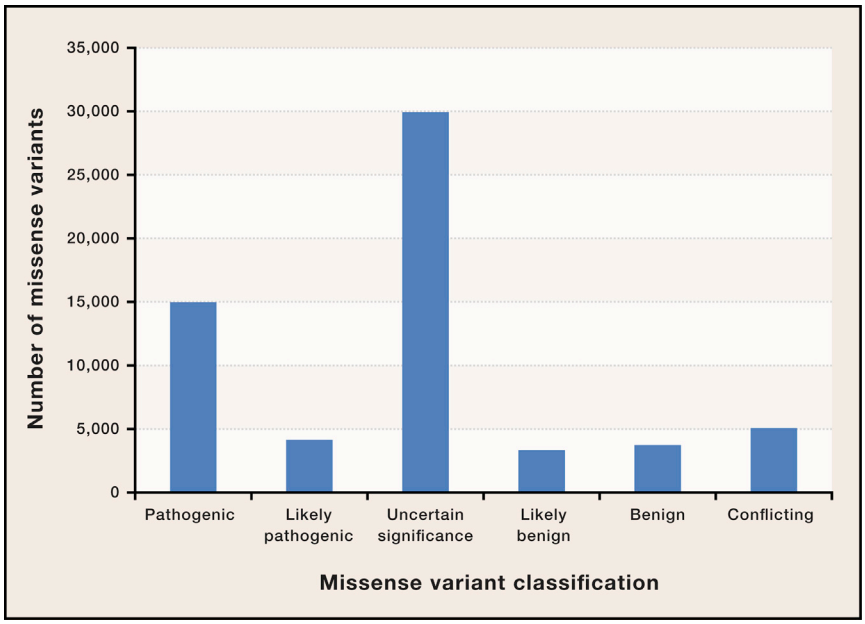

Figure 1. Pathogenicity Assertions
Pathogenicity assertions for 61,169 of 72,472 ClinVar variants as of September 1, 2016. N.B., no pathogenicity assertions provided for 11,303 variants.

into clinical practice. Here, we highlight examples speaking to the value and need for greater basic-clinical integration, describe functional and informatics resources that can facilitate integration, provide recommendations for prioritizing clinically relevant genes for functional investigation, and suggest approaches for promoting these critical interactions.

Linking Genotype to Function

Rare, unexplained, or atypical cases of human disease are increasingly being diagnosed with the help of clinical exome and genome sequencing. Diagnostic yields from sequence data exceeding 25%–40% of previously undiagnosed patients are being reported from genetics laboratories and clinics where diagnoses were previously made at a small fraction of that rate. Diagnostic yield is even higher in infants admitted to high-level neonatal intensive care units, where trio genome sequencing in critically ill newborns with a suspected genetic disorder has yielded definitive diagnoses in over 50% of patients (Willig et al., 2015).

Despite these advances, a major barrier to the interpretation of genomic variants is a lack of functional evidence of pathogenicity. Efforts to classify variants as pathogenic, likely pathogenic, of uncertain significance, likely benign, or benign (Richards et al., 2015) are frequently stymied by the limited functional information on specific variants or the genes harboring them. Even variants clearly demonstrated to be pathogenic might have variable penetrance, underscoring the need for more quantitative, probabilistic approaches to variant interpretation that account for potential complex interactions with other genes, environmental exposures, epigenetic modifications, and other modifying factors. Altogether, this leads to a substantial number of variants being relegated to the category of VUS. Indeed, as of September 1, 2016, 41% of the over 72,000 missense variants in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) are VUSs (Figure 1) (Landrum et al., 2016).

A number of approaches aim to address the question of variant impact. Sophisticated computational algorithms have been developed to predict the functional consequence of genomic variants; however, these predictions are not robust enough on their own for use in a clinical context. A genome-wide atlas of variant impact, using massively parallel functional assays that marry selection for specific protein functions with high-throughput DNA sequencing to quantify activity of protein variants on a massive scale, as has recently been done for the RING domain of *BRCA1* (Starita et al., 2015), would be a valuable complement to existing computational methods.

These approaches show promise for efficient evaluation of the functional relevance of protein-coding variants in known disease genes where the functional assays have rigorous clinical validity. However, similar types of high-throughput assays that assess variants in novel genes and non-coding sequences are needed. Linking such data to patients' phenotypic characteristics and treatment responses could yield additional insights into protein function that could link back to prognosis and treatment, illustrating the tremendous potential for basic and clinical scientists to augment each other's work. Recognizing that assays focused on proteins might not fully capture the functional impact of variation in regulatory elements, high-throughput studies are also needed to identify variants leading to changes in transcriptional output or other consequences on nearby or even distant gene regions. Such studies are likely to increase understanding not only of disease mechanisms, but also of the influence of cell type and developmental stage on disease development and progression.

The armamentarium of variant characterization approaches also includes large, publicly available databases on patterns of human genomic variation in the general population, including the 1000 Genomes Project (http://www.internationalgenome.org/) and the Exome Aggregation Consortium ([ExAC] http://exac.broadinstitute.org/). Now totaling nearly 70,000 publicly available exomes, such databases have transformed clinical variant interpretation by allowing analysts to identify patient variants with frequencies too high to be consistent with pathogenicity and facilitating the assessment of penetrance of disease-causing variants (Minikel et al., 2016).

Integration of Model Organism Phenotype Ontologies with Human Data

Model organisms have long provided key evidence for establishing the role of genes and variants in human disease (Schughart

et al., 2013). The advent of genome editing technologies such as CRISPR/Cas9 has revolutionized the precision, scale, and speed with which new animal and cell models carrying specific genetic variants observed in human diseases can be generated (Ledford, 2015). For some clinical sequencing initiatives, such as the Undiagnosed Diseases Network (https://undiagnosed.hms.harvard.edu/) and the Centers for Mendelian Genomics (http://mendelian.org/), model organism scientists have been closely integrated with clinical teams to develop model organisms carrying potentially pathogenic genome sequence variants and to then characterize the resulting phenotypes (Gahl et al., 2016; Yamamoto et al., 2014). Cell lineages and organoids derived from human induced pluripotent stem cells (iPSCs) also provide powerful approaches to determining function and pathologic impact of disease-related genetic variants—for example, by introducing putative regulatory variants and assessing allelic-specific expression levels in iPSC-derived lineages. A key to the success of such collaborations is access to detailed clinical phenotyping information in patients and their families, data that are often inaccessible to basic investigators.

An emerging problem for the integration of phenotypic knowledge is the specialized and diverse terminology used by different research communities. Although variants in orthologous genes can produce similar phenotypes across model organisms and humans—indeed, this is the major value of model organism studies for understanding human variation—even when similar phenotype terms are used, they may be applied to different biological entities, such as gene, genotype, or allele, by different research communities. Development and use of biomedical ontologies and controlled vocabularies to standardize the representation of data describing gene function, phenotypes, and disease are essential for robust data sharing and assembling evidence across multiple organisms. Development of a minimal set of quantitative phenotypes across phyla would also facilitate scalable annotation of genomes and integration of model organism data.

Some of these problems are being tackled via the Global Alliance for Genomics and Health (GA4GH) Phenotype Exchange Format initiative (http://phenopackets.org/), which is developing data exchange standards and harmonizing annotation methods for phenotype data. Current efforts to align phenotypic profiles in model organisms with human diseases via an ontology-based approach include the Monarch Initiative (https://monarchinitiative.org/), which uses expertly curated annotations from model organism database resources, along with species-neutral ontologies such as the Gene Ontology (http://geneontology.org/) and the Uberon anatomy ontology (https://bioportal.bioontology.org/ontologies/UBERON) to create logical interoperability. Conversely, Gene Weaver (http://geneweaver.org/) infers phenotypic relationships among genes and processes by matching across gene sets. Such approaches allow organism-specific annotations to be computationally evaluated across species. Related efforts to unify phenotypic descriptions across thousands of Mendelian syndromes and across rare and common diseases include the Human Phenotype Ontology ([HPO] http://human-phenotype-ontology.github.io/) and ClinVar's MedGen effort to harmonize major genetic vocabularies (https://www.ncbi.nlm.nih.gov/medgen/). Linking ontology-based clinical and model organism phenotypic profiles and enhancing engagement among providers of model organism database resources and developers of clinical informatics tools and electronic health records (EHRs) could help accelerate the functionalization of VUSs and bridge the basic-clinical divide.

Data Deposition and Integration

Special considerations in data integration include the additional resources needed for deposition of clinical data into databases, which is currently performed by some clinical laboratories and clinicians as an unpaid labor of love. Deposition of data should be the final step in producing a reimbursable clinical report, as it is essential to making the wealth of clinically derived patient data accessible to computational algorithms that will improve both clinical interpretation and biomedical discovery. It should not, however, become an unfunded mandate; rather, the value of data deposition to the subsequent interpretation of all patients' genomes should be recognized, and consideration needs to be given to reimbursing for it accordingly. Improved tools for data deposition, such as ClinVar's Submission Wizard (https://www.ncbi.nlm.nih.gov/home/submit-wizard.shtml), and for standardized collection of phenotyping information, such as Phenotips (https://phenotips.org/), the PatientArchive (www.patientarchive.org), and SimulConsult (http://simulconsult.com/), should be developed and integrated into EHRs. Algorithms that support ontology-based phenotypic comparisons, such as the GA4GH Matchmaker Exchange (www.matchmakerexchange.org), can also be used to identify similar rare or undiagnosed patients globally.

Another key standardization issue related to data deposition relates to the fragmentation of consent and approval processes around submission of data to databases such as ClinVar. Clearly stated guidance on the types of public database submissions that should receive institutional review board approval, whether reviews can be expedited or waived entirely, and what consent (if any) is needed, including specific language, would be enormously helpful in facilitating deposition of these much needed data into public databases.

Improving Basic-Clinical Interactions in Functionalizing Genomic Variants

We propose four broad efforts to help promote rapid translation of clinical genome sequencing efforts into functional understanding at the bench and use in clinical care.

Identify Clinically Relevant Genes as Priorities for Functional Studies

Selecting a consensus set of high-priority genes for functional studies on the basis of potential clinical importance and experimental feasibility that would be regularly updated would provide a key resource for connecting the basic and clinical research communities. These would include genes of known function that are already in common clinical use and for which the clinical impact of newly identified variants could be significant, as well as genes of unknown function strongly implicated in

Box 1. Partial List of Resources that Can Be Used to Align Priorities for Basic Research with Genes and Variants of High Clinical Relevance

The resources listed include genes with known disease associations for which insights into the roles of specific variants would be beneficial and additional resources also applicable to genes implicated in disease but for which research to validate the disease relevance is needed.

KNOWN DISEASE GENES

- Genes recommended for return of secondary findings in clinical genome sequencing (ACMG); https://www.acmg.net/secondaryfindings
- Genes with variants affecting drug response (CPIC), levels A and B; https://cpicpgx.org/
- Genes encoding proteins that are direct targets of cancer therapeutics; https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet
- Genes recognized as clinically actionable (ClinGen); https://clinicalgenome.org/working-groups/actionability/projects-initiatives/actionability-evidence-based-summaries/
- Genes with high clinical validity (ClinGen); https://www.clinicalgenome.org/working-groups/gene-curation/projects-initiatives/clinical-validity-classification/
- Genes with registered genetic tests, as indicator of current clinical testing volume (GTR); https://www.ncbi.nlm.nih.gov/gtr/

KNOWN AND PUTATIVE (NOVEL OR NEWLY IMPLICATED) DISEASE GENES

- Genes strongly implicated in congenital or undiagnosed diseases; https://commonfund.nih.gov/kidsfirst/overview
- Genes with high loss of function intolerance score (ExAC); http://exac.broadinstitute.org/
- Genes annotated with low numbers of missense variants (ClinVar); https://www.ncbi.nlm.nih.gov/clinvar/
- Genes whose expression is altered in disease (GTEx); http://www.gtexportal.org/home/
- Gene families targetable by drugs; http://baderlab.org/Data/RoadsNotTaken
- Genes with an aberrant phenotype in mouse subsequently shown to have a human disease association (MGI); http://www.informatics.jax.org
- Genes predicted to have human disease associations on the basis of data mining of phenotypes across model organisms; https://monarchinitiative.org/

Abbreviations are as follows: ACMG, American College of Medical Genetics and Genomics; ClinGen, Clinical Genome Resource; CPIC, Clinical Pharmacogenetics Implementation; ExAC, Exome Aggregation Consortium; GTEx, Genotype-Tissue Expression project; GTR, Genetic Testing Registry; MGI, Mouse Genome Informatics.

undiagnosed disease, particularly genes under strong evolutionary constraints (Box 1). Criteria for prioritizing genes and variants within such a set could include the volume of clinical use as suggested by the number of clinically available tests (https://www.ncbi.nlm.nih.gov/gtr/), as well as number of conflicting clinical significance reports and measures of intolerance to variation. The feasibility of massively parallel functional assays for such variants could be established based on existing data on cell autonomous function and the length and number of transcripts. Such data are readily accessible from online resources such as GeneCards (http://www.genecards.org) that aggregate gene annotations from many sources.

To exemplify such an analysis, we have here reviewed Known Disease Genes, as identified from the sources described in Box 1, against the criteria described above. Using this approach, the top ten genes in either number of tests or number of conflicts are displayed with their associated metrics in Table 1 and include several clinically important genes, such Lamin A/C (*LMNA*) and *BRCA2*. Large genes in frequent clinical use, such as *BRCA2,* not unexpectedly had higher numbers of missense variants and conflicting pathogenicity reports; correlations of number of tests with either number of missense variants or conflicting reports were $\geq$0.8. All 97 genes identified in guidelines from the American College of Medical Genetics and Genomics (ACMG), the Clinical Pharmacogenetics Implementation Consortium (CPIC), and the Clinical Genome Resource (ClinGen) are displayed in Table S1. It is important to note that the approach proposed here is centered on genes (particularly their coding regions), whereas non-coding variants are associated with disease in ways that are as yet largely unknown. As understanding of non-coding variants grows, they should be incorporated into expanded criteria for prioritizing functional assessments for clinical use.

Develop Larger Reference Variant Databases Linking to Phenotypes

Reference databases have been invaluable for filtering candidate variants, but such resources must continue to grow both in size and (more critically) in ancestral diversity. Yet even ExAC, the largest publicly available human variant frequency dataset, comprises less than 10% of all exome sequences generated worldwide to date—hence the strong emphasis here on data sharing. Substantial power could be gained by creating policies that foster the creation of aggregated resources and supporting aggregation and harmonization activities, including the standardization of data processing and variant-calling pipelines. Populations currently under-represented in frequency databases (essentially all those of non-European ancestry) should be a particular focus, not only for interpreting variants in these populations but also for understanding rare variants shared by diverse populations.

While variant frequency data without phenotypes are useful, their value is vastly enhanced by the addition of phenotypic data. Current public datasets of human variation, aside from ClinVar, are either aggregate frequency resources that do not share individual-level genetic data or are comprised of fully de-identified samples that do not provide individual-level phenotype data, thus missing critical opportunities to link genotype and phenotype. Future resources must work to make phenotype data available in a way

Table 1. Top Ten Genes in Either Number of Tests or Number of Conflicting Pathogenicity Reports, Ordered by Number of Tests.

Gene Symbol	Disease Name or Relevant Drug	Source	No. Registered Tests[a]	No. Conflicting Pathogenicity Reports[b]	No. Missense Variants[c]	LoF Intolerance[c]	Cell Autonomy, Yes/No[d]	No. amino acids[d]	No. Transcripts[d]	Cell Localization[d]	Actionability Score[e]
LMNA	dilated cardiomyopathy 1A (MIM: 115200)	ACMG56	331	97	303	0.99	yes	664	4	nucleus, cytosol	10CB
PTEN	PTEN hamartoma tumor syndrome (MIM: 153480)	ACMG56	263	8	162	0.98	yes	403	5	cytosol	10CC
BRCA2	breast-ovarian cancer, familial 2 (MIM: 612555)	ACMG56	244	732	2,493	0	yes	3,418	3	nucleus	10AA
TP53	Li-Fraumeni syndrome 1 (MIM: 151623)	ACMG56	240	57	303	0.91	yes	393	9	nucleus	10CB
BRCA1	breast-ovarian cancer, familial 1 (MIM: 604370)	ACMG56	222	491	1,411	0	yes	1,863	8	nucleus	10AA
MLH1	Lynch syndrome (MIM: 120435)	ACMG56	222	9	545	0.74	yes	756	9	nucleus	10AA
MSH2	Lynch syndrome (MIM: 120435)	ACMG56	221	10	636	0.87	yes	934	6	nucleus	10AA
FGFR3	fibroblast growth factor receptor	ClinGen	219	3	70	0	yes	806	9	membrane, cytosol	NA
CFTR	ivacaftor	CPIC	210	76	577	0	yes	1,480	11	membrane	NA
MSH6	Lynch syndrome (MIM: 120435)	ACMG56	204	23	743	NA	yes	1,360	9	nucleus	10AA
SCN5A	Brugada syndrome 1 (MIM: 601144); long QT syndrome 3 (MIM: 603830)	ACMG56	202	131	682	1	yes	2,016	5	membrane	10CB
MYH7	familial hypertrophic cardiomyopathy 1 (MIM: 192600)	ACMG56	193	134	738	1	yes	1,935	17	cytosol	10CB
KCNQ1	long QT syndrome 1 (MIM: 192500)	ACMG56	152	195	396	0	yes	676	6	membrane	10CB
MYBPC3	dilated cardiomyopathy 1A (MIM: 115200); familial hypertrophic cardiomyopathy 4 (MIM: 115197)	ACMG56	139	96	411	0	yes	1,274	11	cytosol	10CB
TSC2	tuberous sclerosis 2 (MIM: 613254)	ACMG56	137	121	648	1	yes	1,807	18	cytosol, nucleus	10EB
RYR1	malignant hyperthermia (MIM: 145600); CPIC	ACMG56	102	128	385	0	yes	5,083	11	membrane	10DB
LDLR	familial hypercholesterolemia (MIM: 143890)	ACMG56	58	143	829	0	yes	860	8	membrane, ER, Golgi	11CA

[a]Genetic Testing Registry (GTR).
[b]ClinVar.
[c]Exome Aggregation Consortium (ExAC).
[d]GeneCards.
[e]Clinical Genomics Resource (ClinGen).
Three genes (*BRCA1, BRCA2,* and *LMNA*) are in the top ten for both metrics.

that empowers research without violating participant consent, either by recruiting participants who consent to fully open data sharing or by supporting methods to link phenotype data to aggregated variants in ways that do not compromise participants' confidentiality.

Finally, intuitive and accessible approaches are needed to provide variant frequency and associated phenotype data to clinical users, starting with close integration of such databases with variant interpretation resources such as ClinVar. Over the longer term, user-friendly interfaces are needed that aggregate multiple lines of support for variant interpretation, including previous observations in patients, frequency data from reference databases, and high-throughput functional assays as described above.

Develop and Adopt Standards for Phenotype Description and Data Sharing

A significant advance would be identification of a minimum reportable clinical dataset of phenotypic information for a given condition, with feedback to the reporting clinician as to the completeness of the submitted information, such as that provided through tools such as Phenotips. Such a dataset has recently been defined for somatic variants (Ritter et al., 2016).

The substantial burden on busy clinicians of actually collecting and recording this information, as discussed above, might be addressed by differential reimbursement models based on the additional information provided. Automated extraction of phenotypic information from EHRs through structured field searches and natural language processing is becoming increasingly useful (Denny et al., 2013). Improved methods for patients to provide phenotypic information would also help, including a patient-specific ontology, such as the newly-released patient-specific version of the HPO. Such information, perhaps subject to review and confirmation by a patient's clinician, could not only facilitate interpretation of genome sequencing results but could provide additional phenotyping information pertinent to the clinical evaluation. Commercial entities such as 23andMe have successfully enabled their consumers to provide extensive information; similar resources that are publicly available include openSNP (https://opensnp.org/), ClinGen's GenomeConnect (https://www.clinicalgenome.org/genomeconnect/), and the Sync4Science effort of the Precision Medicine Initiative (http://www.sync4science.org/).

Finding better ways to facilitate the sharing of potentially identifiable clinical data with closely integrated basic science collaborators, including studies of possible policy and regulatory solutions, should be a high priority. Computational phenotypes stripped of protected health information (PHI), for example, can be shared more readily than potentially identifiable complete clinical datasets. Notably, many patients and families with undiagnosed diseases are less concerned about confidentiality in their determined search for a diagnosis and treatment. Asking such patients and families to consent to more open sharing of their clinical data, with full disclosure of the attendant risks, would be one approach to speeding these discoveries. The development of a clear but concise nationally available consent form for data deposition could ensure agreed-upon individual protections are in place while substantially accelerating access to these important resources.

Promote Cross-Disciplinary Understanding and Opportunities for Interaction

The limited experience to date in applying functional genomic and model organism evidence to real-world clinical scenarios, and clinicians' understandable preference for human studies and randomized clinical trial evidence, can present major barriers to incorporating rare genomic variant information in clinical care. Patients with these conditions are simply not available in the numbers needed for such studies, nor are clinical trials feasible for thousands of actionable variants. Guidelines for the types of evidence that should be accepted by clinicians as confirming the clinical relevance of a genomic variant are beginning to be developed (Green et al., 2013; Hunter et al., 2016) and include reproducible and robust "functional studies supportive of a damaging effect on the gene or gene product" (Richards et al., 2015). Such efforts largely speak to the small subset of clinical researchers working in genomics, however; greater emphasis is needed on understanding and meeting expectations of clinicians in general, who may tend to view genomic data that are anything short of definitive as uninformative. This legacy of exceptionalism in expecting genomic diagnostics to be near perfect is not applied to other clinical diagnostic tests—uncertainty pervades all of medicine. Clinicians must still decide whether to act on the information, uncertain or otherwise, so guidance that supports this clinical decision making would be very helpful. Rather than fixating on using variant information to make a definitive diagnosis, it might be useful to consider presence of a variant as a probabilistic biomarker that can improve decision making but only when placed in an overall clinical context, using the Bayesian logic widely employed in most other non-genetic clinical testing.

The substantial attention given here to bridging the gap between basic and clinical genomics should not overshadow the importance of greater integration within these two disciplines. Among basic scientists, better integration across model systems, phenotype ontologies, and high-resolution quantitative phenotypes has been described above as key to more rapid functionalization of unknown variants and genes. Similarly, there is a greater need for interaction between the clinical laboratory performing and interpreting the genome sequence and the clinician diagnosing and treating the patient. The importance of providing detailed phenotyping information to the laboratory to aid in variant interpretation has been described above, but even better would be a two- or even three-way interchange including the patient as a long-term partner. This would enable the clinician to go back to the patient and seek, among other information, key phenotypic manifestations to which the molecular pathologist may have been alerted based on review of available databases.

Other approaches to enhancing basic-clinical integration and understanding each other's questions include fostering opportunities for formal and informal interactions, such as participation in each other's research consortia and scientific meetings. Academic case conferences involving clinical molecular pathologists and domain-specific clinicians, as well

as basic scientists expert in interpreting genomic and functional data, have been particularly successful in using genome sequence data for diagnosis and treatment (Yang et al., 2014). Including informaticists deeply familiar with phenomic ontologies and annotations would provide additional depth. Promoting and expanding data sharing and collaborative initiatives such as those described above will also be of value.

These efforts individually and collectively hold great promise for bringing basic and clinical researchers and clinicians, and indeed researchers from many other relevant disciplines, together to work on mutually relevant questions that will ultimately benefit them both, the scientific community at large, and most importantly, the patients whom we are all committed to serving.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.03.005.

REFERENCES

Cooper, G.M. (2015). Parlez-vous VUS? Genome Res. *25*, 1423–1426.

Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. *31*, 1102–1110.

Gahl, W.A., Mulvihill, J.J., Toro, C., Markello, T.C., Wise, A.L., Ramoni, R.B., Adams, D.R., and Tifft, C.J.; UDN (2016). The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. Mol. Genet. Metab. *117*, 393–400.

Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., et al.; American College of Medical Genetics and Genomics (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. *15*, 565–574.

Hunter, J.E., Irving, S.A., Biesecker, L.G., Buchanan, A., Jensen, B., Lee, K., Martin, C.L., Milko, L., Muessig, K., Niehaus, A.D., et al. (2016). A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. Genet. Med. *18*, 1258–1268.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44*(D1), D862–D868.

Ledford, H. (2015). CRISPR, the disruptor. Nature *522*, 20–24.

Minikel, E.V., Vallabh, S.M., Lek, M., Estrada, K., Samocha, K.E., Sathirapongsasuti, J.F., McLean, C.Y., Tung, J.Y., Yu, L.P., Gambetti, P., et al.; Exome Aggregation Consortium (ExAC) (2016). Quantifying prion disease penetrance using large population control cohorts. Sci. Transl. Med. *8*, 322ra9.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.

Ritter, D.I., Roychowdhury, S., Roy, A., Rao, S., Landrum, M.J., Sonkin, D., Shekar, M., Davis, C.F., Hart, R.K., Micheel, C., et al.; ClinGen Somatic Cancer Working Group (2016). Somatic cancer variant curation and harmonization through consensus minimum variant level data. Genome Med. *8*, 117.

Schughart, K., Libert, C., and Kas, M.J.; SYSGENET consortium (2013). Controlling complexity: the clinical relevance of mouse complex genetics. Eur J Hum Genet. *21*, 1191–1196.

Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics *200*, 413–422.

Willig, L.K., Petrikin, J.E., Smith, L.D., Saunders, C.J., Thiffault, I., Miller, N.A., Soden, S.E., Cakici, J.A., Herd, S.M., Twist, G., et al. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. Lancet Respir. Med. *3*, 377–387.

Yamamoto, S., Jaiswal, M., Charng, W.L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell *159*, 200–214.

Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. JAMA *312*, 1870–1879.