Large-Scale Medical Resequencing for X-Linked Mental Retardation

Yiping Shen,1,2,3 Bai-Lin Wu,2,4 and James F. Gusella1,5*

Severe intellectual disability (ID),6 commonly referred to as mental retardation (MR), comprises a large collection of clinical conditions whose associated phenotypes include substantially below-average intelligence test scores and limited abilities in socially adaptive behaviors, such as communication, self-care, social interaction, and school functioning. ID/MR affects 1%–3% of the worldwide population with variable severity, is a frequent cause of pediatric, genetic, neurologic, and developmental medicine referrals, and poses huge social and economic burdens. Genetic factors, such as trisomy 21 in Down syndrome, and environmental risk factors, such as fetal alcohol syndrome, contribute to the pathogenesis of ID/MR. Although the genetic causes are heterogeneous and complex, a 30%–40% excess of males vs females and the existence in many families of an evident X-linked inheritance pattern suggest a major role for inactivating mutations in X-chromosome genes. The notion is that these mutations cause ID/MR predominantly in males, in whom, unlike females, there is no second X chromosome present to complement the defect and thereby inhibit its phenotypic expression. Consequently, a variety of molecular genetic strategies have been applied to the X chromosome, including candidate gene analysis, linkage mapping followed by location cloning, and precise delineation of chromosomal anomalies (balanced translocation, inversion, and microdeletion/duplication). These efforts have led to the discovery of a trinucleotide expansion in the 5’ untranslated region of the FMR17 (fragile X mental retardation 1) gene as the cause of fragile X syndrome, the most frequent form of X-linked ID/MR, and mutations in approximately 90 other genes in rarer families (1, 2). Because known genes account for less than half of the >200 delineated X-linked ID/MR conditions (http://xlmr.interfree.it/home.htm), many novel X-linked genetic factors have yet to be identified.

To address this issue, Tarpey et al. recently undertook a “drown the pond” approach in which they systematically examined the sequences of coding exons for 718 X-chromosome genes in 208 X-linked ID/MR individuals prescreened for known X-linked ID/MR gene mutations (3). Although this project successfully identified a number of new ID/MR-associated genes, it demonstrated the challenge of the direct-sequencing approach and presaged, for both the X chromosome and autosomes, the complexity of interpreting the genomic sequence data that are expected to accumulate rapidly with the advent of next-generation sequencing technologies (4).

The global X-chromosome exon-sequencing approach used by Tarpey et al. offers a strategy unbiased by assumptions concerning the nature of the genes involved in ID/MR. The positive outcome of the screen was the identification of 9 X-linked ID/MR genes: API52 (adaptor-related protein complex 1, sigma 2 subunit), CUL4B (cullin 4B), BRWD3 (bromodomain and WD repeat domain containing 3), UPE3B (UPF3 regulator of nonsense transcripts homolog B (yeast)), ZDHHC9 (zinc finger, DHHC-type containing 9), SLC9A6 (solute carrier family 9 (sodium/hydrogen exchanger), member 6), SYP (sypaptophysin), ZNF711 (zinc finger protein 711), and CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)). These results were based on both the occurrence of one or more truncating mutations in cases but not controls and confirmation by family-based segregation with the ID phenotype. In some cases, targeted follow-up sequencing was carried out in 914 additional individuals and 1129 unaffected control individuals. Although these findings raised to approximately 11% the proportion of X-chromosome genes in which inactivating mutations are associated with ID/MR, they left nearly 75% of the X-linked ID/MR families studied without an identified mutation. Although some of these lesions could lie in gene-coding sequences on the X chromosome not assessed in this study (approximately 35% of

1 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2 Department of Laboratory Medicine, Children’s Hospital Boston, Boston, MA; Departments of 3 Neurology, 4 Pathology, and 5 Genetics, Harvard Medical School, Boston, MA.

* Address correspondence to this author at: Massachusetts General Hospital/ Harvard Medical School, 185 Cambridge St., Boston, MA 02114. Fax 617-726-5735; e-mail gusella@helix.mgh.harvard.edu.

Received October 6, 2009; accepted December 11, 2009.

Previously published online at DOI: 10.1373/clinchem.2009.135020

Nonstandard abbreviations: ID, intellectual disability; MR, mental retardation.

1 Human genes: FMR1, fragile X mental retardation 1; API52, adaptor-related protein complex 1, sigma 2 subunit; CUL4B, cullin 4B; BRWD3, bromodomain and WD repeat domain containing 3; UPE3B, UPF3 regulator of nonsense transcripts homolog B (yeast); ZDHHC9, zinc finger, DHHC-type containing 9; SLC9A6, solute carrier family 9 (sodium/hydrogen exchanger), member 6; SYP, synaptophysin; ZNF711, zinc finger protein 711; CASK, calcium/calmodulin-dependent serine protein kinase (MAGUK family).
the total) or in autosomal genes with a male-limited phenotype that in small families mimics X-linked inheritance or they could represent gross genomic anomalies not detectable by this sequencing approach, it is unlikely that such lesions could account for the majority of the genetic defects not yet found. Consequently, it is likely that many X-linked ID/MR mutations lie outside of coding exons and affect regulatory elements or sequences encoding non–protein-coding RNAs.

An obvious approach to identify the “missing” genetic factors is to sequence the entire X chromosome, a task now feasible with next-generation massively parallel sequencing methods that dramatically increase the yield and decrease the cost of DNA sequence determination. Indeed, these technological advances raise the prospect of advancing to a fully unbiased strategy to discover all genetic factors that contribute to ID/MR, including those on the autosomes; however, the study of Tarpey et al. offers a sobering indication that even the power of sequencing the complete genome will not by itself resolve the issue of genetic factors in ID/MR. In their scan of 718 genes, these investigators identified truncating alterations in 19 genes (representing >1% of all X-chromosome genes) that proved not to be associated with ID/MR, because these variants were also found in unaffected controls or did not cosegregate with the ID/MR phenotype in the family in question. Thus, even the obvious functional consequence of the lesion was not sufficient to conclude that it caused ID/MR. Instead, family-based segregation analysis and examination of population-based prevalence were critical for data interpretation. The same argument applies to nontruncating variants, the precise functional importance of which is unknown. This category of variants was also detected in large numbers in this study, and their proportion among variants detected is likely to increase as more noncoding sequences are examined. Furthermore, interpreting rare (“private”) genetic variants on the autosomes promises to be even more difficult, given that testing for familial segregation is often impossible.

The implications of studies such as that of Tarpey et al. for molecular diagnostics laboratories and their adherence to American College of Medical Genetics guidelines are profound (5). Increasingly, diagnostic testing for complex disorders such as ID/MR requires an approach based on a panel of genes rather than a single gene to provide the most comprehensive information possible concerning genetic etiology. In parallel, improvements in microarray technologies have provided the capacity to scan the entire genome in an unbiased manner for copy number alterations that may be associated with disease, including those seen in ID/MR. The genetic heterogeneity of complex, common disorders such as ID/MR and the high prevalence of nonrecurring private sequence variants suggest that although full sequencing of a large panel of genes is required, the interpretation of individual variants, even de novo alterations, may often remain problematic, as it is often for individual copy number variants detected in array comparative-hybridization or single-nucleotide polymorphism microarray studies. Consequently, to successfully respond to this changing diagnostic paradigm, molecular diagnostics laboratories will require substantial enhancements in infrastructure, both to enable the efficient collection of large volumes of DNA sequencing and genomic microarray data on clinical test samples and to support the interpretation of the results. The latter in particular calls for a dramatic improvement in bioinformatics capabilities within molecular diagnostics laboratories and a concerted effort by the genetics community to assemble and maintain dynamic databases of cumulative sequence and copy number findings (from both clinical and unaffected control samples) to permit the most accurate interpretation possible of the variants observed.

In the case of ID/MR, active ascertainment and inclusion of phenotypic correlates in the databasing effort may also help in classifying patients and interpreting molecular findings, thereby improving the clinical utility of the genetic test panel. ID/MR is a condition with very diverse phenotypes that is classified as syndromic when there is an accompanying medical condition and as nonsyndromic when ID/MR is the only feature. Such recognized classifications can hasten the clinical recognition of families with specific types of ID/MR and can subsequently facilitate molecular genetic studies. The boundary between syndromic and nonsyndromic ID/MR is not clear in many cases, however, and both syndromic and nonsyndromic ID/MR can be caused by defects in the same gene. In many situations, identifying a genetic defect in a specific gene provides the ultimate evidence for a clinical diagnosis of a patient with a specific ID/MR disorder; therefore, access to a comprehensive cumulative database of both phenotypic and molecular characteristics of all prior cases ascertained and tested will continuously improve the quality of test interpretation. Such a resource would also enable the molecular diagnostics laboratory to provide accurate guidance to researchers, thereby enhancing their efforts to understand the mechanisms involved in ID/MR and to develop successful interventions based on data obtained directly from individuals with the disorder. Current human gene mutation databases, such as the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff (http://www.hgmd.cf.ac.uk) and the Leiden Open Variation Database for mental retardation (http://www.LOVD.nl/MR), are very useful resources, but they are limited by their lack of phenotypic information. What is needed is
a comprehensive mutation database that includes both
the genotype and the detailed phenotype, similar to
DECIPHER (Database of Chromosome Imbalance and
Phenotype in Humans using Ensemble Resources; https://decipher.sanger.ac.uk) for copy number vari-

tant mutations.

Although X-linked ID/MR represents a more trac-
table target for molecular research than autosomal ID/
MR, it can serve as an informative model for this and
other complex disorders in which genetic factors play a
major role. The lessons learned from attempts to deline-
ate the full spectrum of genetic causes of X-linked
ID/MR will be invaluable in tackling other complex
neurodevelopmental/neuropsychiatric disorders, such
as autism spectrum disorders, attention deficit disor-
der, and schizophrenia. The study of Tarpey et al. has
revealed the feasibility of interrogating a very large
panel of genes associated with a common phenotype—
not a trivial task given that further technical improve-
ments will be required to make it more reliable and
informative in the setting of the molecular diagnostics
laboratory. Indeed, a number of strategies from the re-
search laboratory are now being adapted and tested in
molecular diagnostics laboratories, with the goal of
rapid, efficient high-throughput generation of reliable
DNA sequence from large gene regions or gene panels.
Although the pace of innovation in the DNA-
sequencing arena promises to solve the technical issues,
the major issue of interpretation of the findings will
remain, because it is inherent to the genetic architec-
ture of ID/MR and likely of many other developmental
conditions. The burden of accurate interpretation of
the molecular data will require a radical improvement
in the infrastructure of molecular diagnostics laborato-
ries, including an increasing emphasis on computa-
tional capabilities in population genetics, construc-
tion and continuous updating of nonpathologic refer-
ce and clinical databases that include both phenotypic
and molecular data, and provisions for efficient data storage, data retrieval, and bioinformatics and sta-

tistical analyses. None of these requirements is
beyond the realm of possibility, but they will require a
cooperative effort of the genetic diagnostics commu-
nity, backed by the will (and the dollars) of the health-
care system.

Author Contributions: All authors confirmed they have contributed to
the intellectual content of this paper and have met the following 3 re-
quirements: (a) significant contributions to the conception and design,
aquisition of data, or analysis and interpretation of data; (b) drafting
or revising the article for intellectual content; and (c) final approval of
the published article.

Authors’ Disclosures of Potential Conflicts of Interest: Upon
manuscript submission, all authors completed the Disclosures of Poten-
tial Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: None declared.
Consultant or Advisory Role: J. Gusella, Quest Diagnostics.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the
design of study, choice of enrolled patients, review and interpreta-
tion of data, or preparation or approval of manuscript.

References

422–34.
landscape of intellectual disability arising from
S, Hardy C, et al. A systematic, large-scale rese-
quencing screen of X-chromosome coding exons in
4. Tucker T, Marra M, Friedman JM. Massively paral-
ellel sequencing: the next big thing in genetic med-
5. Richards CS, Bale S, Bellissimo DB, Das S, Grody
WW, Hegde MR, et al. ACMG recommendations
for standards for interpretation and reporting of