EDITORIAL

A 2.5-year snapshot of Mendelian discovery

Benjamin D. Solomon[1,2,3], Teresa Lee[2], Anh-Dao Nguyen[4] & Tyra G. Wolfsberg[4]

[1]Division of Medical Genomics, Inova Translational Medicine Institute, Inova Health Systems, Falls Church, Virginia 22042
[2]Department of Pediatrics, Inova Children's Hospital, Inova Health Systems, Falls Church, Virginia 22042
[3]Department of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298
[4]Computational and Statistical Genomics Branch, National Human Genome Research Institute, Bethesda, Maryland 20892

Introduction

New sequencing technologies are becoming increasingly available in a variety of contexts. These techniques include whole-exome (we refer to this as "exome" sequencing) and whole-genome sequencing (Biesecker and Green 2014). These types of genomic sequencing have impacted both research and clinical practice. Genomic sequencing has led to the discovery of novel genetic etiologies of disease and has also improved the ability to diagnose patients with subtle or atypical presentations of genetic conditions (especially those affected by relatively rare disorders) by allowing simultaneous interrogation of many loci (Boycott et al. 2013; Yang et al. 2013, 2014; Taylor et al. 2015).

To quantify genetic progress and the impact of these technologies in understanding the causes of Mendelian disorders, we analyzed methods of discovery in the last ~2.5 years.

Materials and Methods

We reviewed all newly described Mendelian disease genes in the ~2.5-year period (April 30, 2013–November 30, 2015) following the initial public dissemination of the Clinical Genomic Database (CGD) (Solomon et al. 2013), a freely available web-based resource that focuses on the clinical sequelae and management of genetic disorders (the CGD, which is regularly updated to keep pace with genetic knowledge, is available at: http://research.nhgri.nih.gov/CGD/). Regarding relevant literature, as there is frequently a gap between manuscript acceptance, electronic, and final publication, we attempted to include only those genes/conditions in the ~2.5 year period that were published (available in PubMed) and subsequently included in both the CGD and Online Mendelian Inheritance in Man (OMIM, available at http://www.omim.org) within that ~2.5 year interval. We recognize this is imperfect for several reasons. These reasons include the lag between publication and incorporation in these databases, and the fact that the databases do not completely capture the literature and are themselves being continually being refined and updated. However, with these caveats in terms of potential inaccuracies, the trends that we reveal are overall interesting.

For each article, we determined how the discovery was made (e.g., through whole-genome sequencing alone, homozygosity mapping and exome sequencing, candidate gene studies, or a combination of possibilities). If the discovery involved findings through a previous publication on the same families or condition (such as linkage analysis implicating specific loci), we included that previous method as part of the discovery process. In the example given, this would be considered "exome + linkage." We also determined if the discovery was made through analysis of a single individual; a single family; or a single individual or family followed by the identification of additional mutation-positive individuals through studies of a larger cohort. Finally, we investigated whether additional bench-based studies, aimed at providing understanding and evidence beyond clinical and bioinformatics results, were included in the investigation. We did not include newly reported conditions allelic to previously described conditions. Regarding this latter exclusion criterion, though many genes have clinically distinct allelic disorders, we wished to be maximally conservative in order to avoid any controversy arising from similar disorders that may represent the spectrum of a single disease entity.

Results

For newly discovered genetic causes, 445 new genes were identified in the 2.5-year period in the defined intervals (see Fig. 1 and Table S1 for details). A number of genes

A. Total
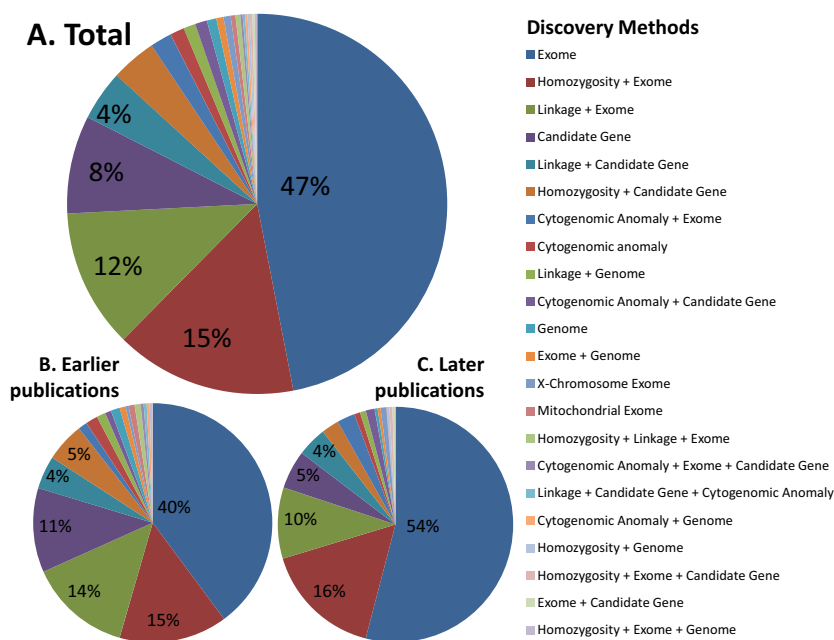
Figure 1. Discovery methods related to the 445 genes (representing 492 studies) newly identified as involved in human Mendelian disease within the selected ~2.5-year period. The top chart (A) shows the overall breakdown of methods within the 2.5-year period; (B) shows the rough first chronological half of publications (sorted by PubMed IDs), while (C) shows the rough second chronological half. There was significantly more use of genomic sequencing, both alone and in conjunction with other techniques, in the second half of the period.

Discovery Methods

- Exome
- Homozygosity + Exome
- Linkage + Exome
- Candidate Gene
- Linkage + Candidate Gene
- Homozygosity + Candidate Gene
- Cytogenomic Anomaly + Exome
- Cytogenomic anomaly
- Linkage + Genome
- Cytogenomic Anomaly + Candidate Gene
- Genome
- Exome + Genome
- X-Chromosome Exome
- Mitochondrial Exome
- Homozygosity + Linkage + Exome
- Cytogenomic Anomaly + Exome + Candidate Gene
- Linkage + Candidate Gene + Cytogenomic Anomaly
- Cytogenomic Anomaly + Genome
- Homozygosity + Genome
- Homozygosity + Exome + Candidate Gene
- Exome + Candidate Gene
- Homozygosity + Exome + Genome

B. Earlier publications

C. Later publications

were identified by apparently independent studies with separate but simultaneous publications. Methods from each such independent study were counted separately here, such that 492 individual studies are represented. Of the 492 studies, 394 (80%) used some type of genomic sequencing (including exome or genome sequencing but not X-chromosome exome or mitochondrial exome). Three hundred and seventy-nine (77%) used exome sequencing, 11 (2%) used genome sequencing, and four (1%) used both exome and genome sequencing in the same study. Two hundred and thirty-one (47%) used only exome sequencing (without other investigations such as homozygosity mapping) for gene identification; 238 (48%) used only exome and/or genome sequencing. Another 156 (32%) used exome or genome sequencing in addition to other techniques, such as homozygosity mapping. Forty-one (8%) used a traditional candidate gene approach without other methods, though 46 (10%) used a candidate gene approach with methods other than exome or genome sequencing. The remaining 11 (2%) used other methods, such as cytogenomic methods alone, X-chromosome exome sequencing, or mitochondrial sequencing (see Table S1).

Eighteen (4%) of the overall studies investigated a single patient only; 114 (23%) studied multiple members of a single family; 93 (19%) were initially done on a single patient or family but then investigated a larger cohort based on these initial findings and identified and reported additional mutation-positive individuals. The remaining 267 (54%) were reported as studying multiple patients or families simultaneously.

We would predict that over time, the use of genomic sequencing methods will eclipse other methods of identifying disease genes. To detect if there has been a shift in methods, even over the short ~2.5 year period of this study, we sorted the 492 publications by PubMed identification number (PMID) and split the publications into two groups. We recognize that the PMID identifiers do not precisely capture chronology. Nevertheless, we note that studies in the second group, which represents manuscripts published more recently, made significantly more use of genomic sequencing overall (210 vs. 184 studies, $P = 0.0046$ by Fisher's exact test) and genomic sequencing alone (135 vs. 103 studies, $P = 0.0051$ by Fisher's exact test) than did those in the first.

Of the 492 studies, 359 (73%) included some type of laboratory-based assay in addition to standard clinical work-up and genomic sequencing and analysis. However, we intentionally did not further analyze the type of cellular or functional analyses that were done as part of the paper – we viewed this assignment as difficult and potentially unhelpful to interpret, as this depended on the availability of previous knowledge about a particular gene's function (e.g., through an animal model), the specific mutation in question (e.g., a truncating vs. missense mutation), may have been aimed at understanding the biological implications of the genetic pathways involved rather than or in addition to the question of mutation pathogenicity, and because so many different possible methods might have been used (MacArthur et al. 2014).

Discussion

Our analyses show that exome sequencing currently accounts for the vast majority of causal gene discovery. Almost half of all the discoveries investigated here occurred through exome sequencing alone. This is anticipated to shift to genome sequencing as affordability and accuracy of sequencing continue to improve (Hayden 2014). Additionally, our analyses show a statistically significant increase in the use of exome or genome sequencing even within the short time period analyzed. Again, though we admit to potential inaccuracies as the background databases shift, we feel that the trends described here are illustrative.

In conclusion, new sequencing technologies are resulting in a dramatic change in the discovery of the causes of disease. These discoveries can be quickly translated into clinical care – patients affected with many conditions now have a better chance of explanations based on genetic testing. Further, finding the molecular etiology of a condition may then in turn yield more tailored and overall better management (Solomon et al. 2013; Soden et al. 2014; Khromykh and Solomon 2015; Solomon 2015; Willig et al. 2015). It will be interesting to perform similar analyses in the future as new techniques emerge and as the body of knowledge of the causes of human disease continues to grow and evolve.

Acknowledgments

Conflict of Interest

None of the authors have disclosures or conflicts of interest.

References

Biesecker, L. G., and R. C. Green. 2014. Diagnostic clinical genome and exome sequencing. N. Engl. J.Med. 370: 2418–2425.

Boycott, K. M., M. R. Vanstone, D. E. Bulman, and A. E. MacKenzie. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat. Rev. Genet. 14:681–691.

Hayden, E. C. 2014. Technology: the $1,000 genome. Nature 507:294–295.

Khromykh, A., and B. D. Solomon. 2015. The benefits of whole-genome sequencing now and in the future. Mol. Syndromol. 6:108–109.

MacArthur, D. G., T. A. Manolio, D. P. Dimmock, H. L. Rehm, J. Shendure, G. R. Abecasis, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. Nature 508:469–476.

Soden, S. E., C. J. Saunders, L. K. Willig, E. G. Farrow, L. D. Smith, J. E. Petrikin, et al. 2014. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Sci. Transl. Med. 6:265ra168.

Solomon, B. D. 2015. Genomic sequencing and the impact of molecular diagnosis on patient care. Mol. Syndromol. 6:4–6.

Solomon, B. D., A. D. Nguyen, K. A. Bear, and T. G. Wolfsberg. 2013. Clinical genomic database. Proc. Natl Acad. Sci. USA 110:9851–9855.

Taylor, J. C., H. C. Martin, S. Lise, J. Broxholme, J. B. Cazier, A. Rimmer, et al. 2015. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat. Genet. 47:717–726.

Willig, L. K., J. E. Petrikin, L. D. Smith, C. J. Saunders, I. Thiffault, N. A. Miller, 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.

Yang, Y., D. M. Muzny, J. G. Reid, M. N. Bainbridge, A. Willis, P. A. Ward, et al. 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J.Med. 369:1502–1511.

Yang, Y., D. M. Muzny, F. Xia, Z. Niu, R. Person, Y. Ding, et al. 2014. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 312:1870–1879.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Spreadsheet of genes, conditions, and articles included in this analysis.